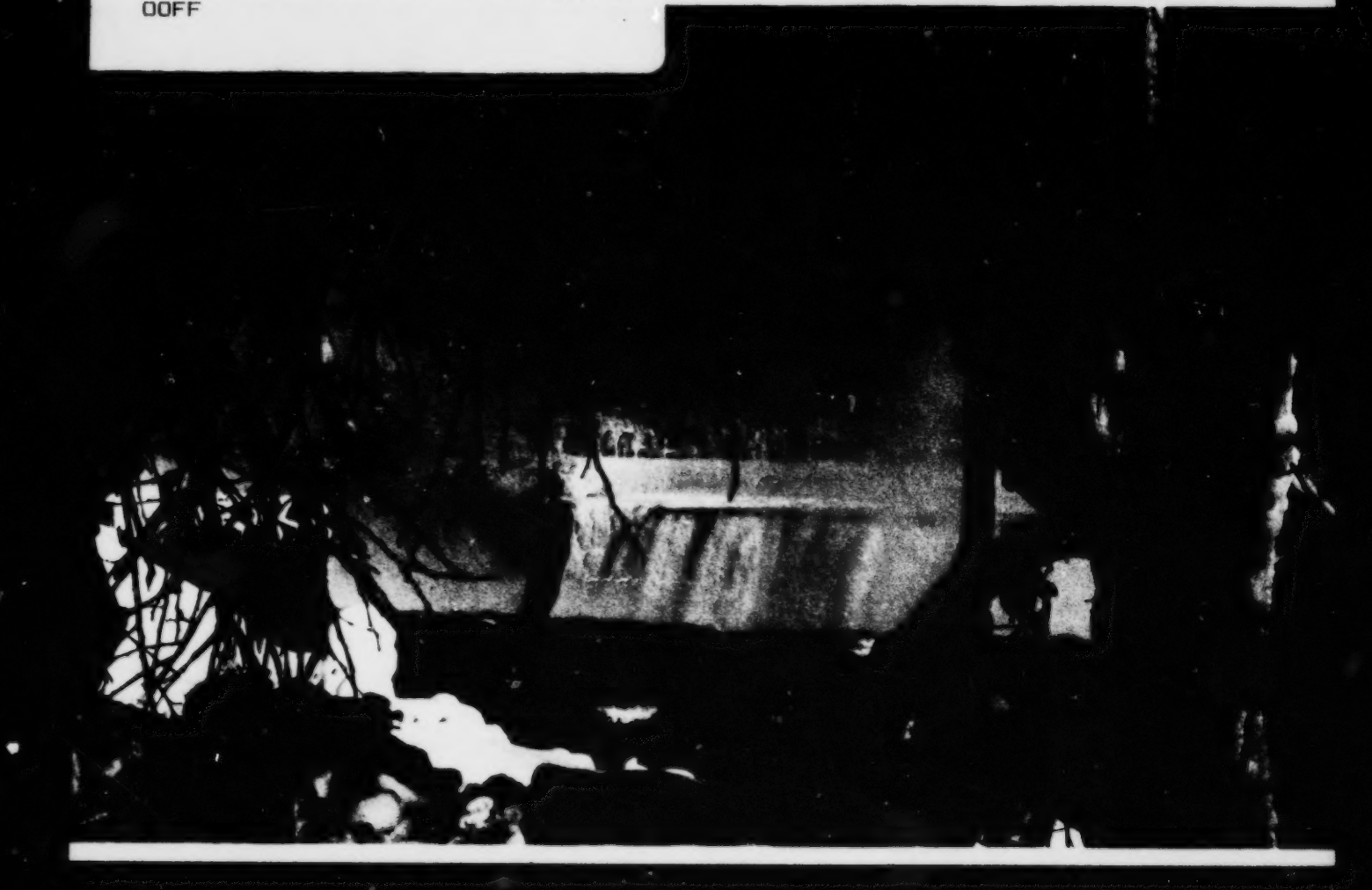
Environment CANADA Environnement
0026272J VOL ISS 136 1985
EXC # 001
GB 707 C3354
SERIE SCIENTIFIQUE (CANADA. DIRECTI
ON GENERALE DES EAUX INTERIEURES)
OOFF



86-01759/

Environnement Canada
Environment Canada

# Évaluation statistique du programme de surveillance des Grands lacs, 1966-1981, lac Érié

Préparé par A.H. El-Shaarawi

ÉTUDE N° 136, SÉRIE SCIENTIFIQUE

*INSTITUT NATIONAL DE RECHERCHE SUR LES EAUX*
*DIRECTION GÉNÉRALE DES EAUX INTÉRIEURES*
*CENTRE CANADIEN DES EAUX INTÉRIEURES*
*BURLINGTON (ONTARIO) 1985*

*(Available in English on request)*

© Ministre des Approvisionnements et Services Canada 1985

N° de cat. En 36-502/136F

ISBN 0-662-93536-5

# Table des matières


| | Page |
|---|---|
| RESUME | xix |
| ABSTRACT | xxi |
| REMERCIEMENTS | xxiii |
| 1. INTRODUCTION | 1 |
| 2. RETROSPECTIVE DU LAC ERIE | 3 |
| Premiers travaux sur les gaz dissous | 4 |
| Variables physiques, biologiques et chimiques régissant la perte de l'oxygène dissous | 6 |
| Etat actuel du lac Erié | 9 |
| Conclusions | 19 |
| 3. VARIATIONS TEMPORELLES DANS LE LAC ERIE | 28 |
| Introduction | 28 |
| Sources des données | 30 |
| Méthodes statistiques | 31 |
| Réduction de la variabilité spatiale | 31 |
| Détermination du cycle saisonnier | 33 |
| Détection des tendances | 34 |
| Niveau de l'eau | 35 |
| Un modèle pour le niveau de l'eau | 35 |
| Le modèle | 36 |
| Variations saisonnières | 39 |
| Débit de la rivière Niagara | 39 |
| Température de l'air | 40 |
| Température de l'eau | 43 |
| Cycle saisonnier de la température de surface de l'eau | 43 |
| Analyse statistique des données de température de surface | 44 |
| Variations d'une année à l'autre de la température de surface de l'eau | 46 |
| Variation verticale de la température | 47 |
| Variabilité d'une année à l'autre de la température de l'hypolimnion (bassin central) | 48 |
| Chlorophylle *a* | 49 |
| Bassin ouest | 49 |
| Bassin central | 51 |
| Bassin est | 51 |

Page

Phosphore total (PT)..... 52
Bassin ouest..... 52
Bassin central..... 53
Bassin est..... 53
Phosphore réactif soluble (PRS)..... 53
Nitrate et nitrite filtrés ($N\text{-}NO_3NO_2$)..... 54
Ammoniac ($NH_3$)..... 55
Profondeur au disque de Secchi..... 56
Variabilité d'une année à l'autre des profondeurs au disque de Secchi..... 57
Turbidité..... 57
Analyse des résultats de l'ajustement du modèle 3.4 aux données logarithmiques de turbidité..... 58
Variabilité de la turbidité d'une année à l'autre... 58
Relation phosphore-chlorophylle a..... 58
Rapport entre les teneurs en ammoniac et celles en nitrite et nitrate..... 59

4. VARIABILITE SPATIALE ET TEMPORELLE DE L'OXYGENE DISSOUS DANS LE LAC ERIE..... 103
Résumé..... 103
Introduction..... 104
Données et méthodes d'analyse..... 107
Méthode de regroupement en grappes..... 107
Application de la méthode de regroupement en grappes..... 111
Taux d'appauvrissement - Calculs et tests..... 112
Résultats et discussion..... 115

5. MODELE STATISTIQUE DE L'OXYGENE DISSOUS DANS LE BASSIN CENTRAL DU LAC ERIE..... 131
Introduction..... 131
Elaboration du modèle..... 133
Prévision du taux d'appauvrissement et tendance des séries chronologiques..... 140
Prévision de l'anoxie et de l'hypolimnion..... 142
Conclusion..... 144

6. DISTRIBUTION SPATIALE ET TEMPORELLE DES CONCENTRATIONS DE COLIFORMES TOTAUX ET FECAUX, 1966-1970..... 150
Introduction..... 150
Méthodes statistiques..... 152
Distributions de fréquences empiriques..... 152
Distributions de probabilités des concentrations bactériennes..... 153
Ajustement des distributions de probabilités et estimation des paramètres..... 154

Page

Méthodes par regroupement en zones..... 161
Un exemple: l'analyse des données de 1968 sur les coliformes totaux..... 164
Inférence relative au rapport des concentrations bactériennes..... 170
Analyse des données sur les coliformes totaux..... 173
Analyse des données sur les coliformes fécaux..... 182
Résumé et discussion..... 189

7. STRATEGIE D'ECHANTILLONNAGE POUR FUTURE CUEILLETTE DE DONNEES..... 234
Introduction..... 234
Illustration des principes de l'échantillonnage..... 236
Stratégie d'échantillonnage d'un lac divisé en zones... 239
Applications..... 240
Généralisation..... 243
Stratégie d'échantillonnage pour une concentration à fonction continue..... 245
Applications..... 246

REFERENCES..... 253

ANNEXE. Examen de modèles de régression linéaire par analyse des résidus..... 265

# Tableaux


| | | Page |
|---|---|---|
| Tableau 1. | Données relatives aux Grands lacs..... | 22 |
| Tableau 2. | Rapport oligotrophie/eutrophie..... | 22 |
| Tableau 3. | Caractéristiques moyennes de l'hypolimnion du bassin central du lac Erié, de 1973 à 1980..... | 23 |
| Tableau 4. | Caractéristiques des teneurs en oxygène dissous du bassin central du lac Erié, de 1930 à 1980... | 23 |
| Tableau 5. | Valeurs de EC pour les zones géographiques du lac Erié..... | 61 |
| Tableau 6. | Résumé des étapes de modélisation du niveau d'eau. | 62 |
| Tableau 7. | Température de l'air du bassin central: moyennes mensuelles..... | 62 |
| Tableau 8. | Résidus et sommes cumulatives des résidus pour la température de l'air du bassin central..... | 63 |
| Tableau 9. | Résumé de la structure de la température de l'air du bassin central, 1967 à 1978..... | 63 |
| Tableau 10. | Résultats de l'ajustement du modèle 3.4 aux températures de surface du lac Erié, 1967 à 1978..... | 64 |
| Tableau 11. | Valeurs de la statistique F pour le test d'égalité des cycles annuels de température de l'eau..... | 65 |
| Tableau 12. | Température moyenne de l'hypolimnion, température au jour 200 du calendrier et pente et ordonnée à l'origine de la droite de régression, 1967 à 1980..... | 65 |
| Tableau 13. | Valeurs estimées des paramètres du modèle 3.4 pour la chlorophylle *a* (bassin ouest)..... | 65 |
| Tableau 14. | Valeurs de $F_1$, $F_2$ et $F_0$ pour la chlorophylle *a* (bassin ouest)..... | 66 |

Page

Tableau 15. Valeurs estimées des paramètres du modèle 3.6 pour la chlorophylle *a* (bassin ouest)..... 66

Tableau 16. Valeurs estimées des paramètres du modèle 3.4 pour la chlorophylle *a* (bassin central)..... 66

Tableau 17. Valeurs de $F_1$, $F_2$ et $F_0$ pour la chlorophylle *a* (bassin central)..... 66

Tableau 18. Valeurs estimées des paramètres du modèle 3.6 pour la chlorophylle *a* (bassin central)..... 67

Tableau 19. Valeurs estimées des paramètres du modèle 3.4 pour la chlorophylle *a* (bassin est)..... 67

Tableau 20. Valeurs de $F_1$, $F_2$ et $F_0$ pour la chlorophylle *a* (bassin est)..... 67

Tableau 21. Valeurs estimées des paramètres du modèle 3.6 pour la chlorophylle *a* (bassin est)..... 67

Tableau 22. Valeurs estimées des paramètres du modèle 3.4 pour le PT (bassin ouest)..... 68

Tableau 23. Valeurs de $F_1$, $F_2$ et $f_0$ pour le PT (bassin ouest)..... 68

Tableau 24. Valeurs estimées des paramètres du modèle 3.6 pour le PT (bassin ouest)..... 68

Tableau 25. Valeurs estimées des paramètres du modèle 3.4 pour le PT (bassin central)..... 69

Tableau 26. Valeurs de $F_1$, $F_2$ et $F_0$ pour le PT (bassin est)..... 69

Tableau 27. Valeurs estimées des paramètres du modèle 3.6 pour le PT (bassin central)..... 69

Tableau 28. Valeurs estimées des paramètres du modèle 3.4 pour le PT (bassin est)..... 70

Tableau 29. Valeurs de $F_1$, $F_2$ et $F_0$ pour le PT (bassin est)..... 70

Tableau 30. Valeurs estimées des paramètres du modèle 3.6 pour le PT (bassin est)..... 70

Tableau 31. Valeurs de $F_1$, $F_2$ et $F_0$ pour le PRS (les trois bassins)..... 70

Page

Tableau 32. Valeurs estimées des paramètres du modèle 3.6 pour le PRS (les trois bassins)..... 71

Tableau 33. Valeurs estimées de $\alpha_t$ et de $R_t$ pour le $N\text{-}NO_3NO_2$ (les trois bassins)..... 71

Tableau 34. Valeurs de $F_1$, $F_2$ et $F_0$ pour le $N\text{-}NO_3NO_2$ (les trois bassins)..... 71

Tableau 35. Valeurs estimées de $\alpha_t$ et de $R_t$ pour le $N\text{-}NO_3NO_2$ (les trois bassins)..... 72

Tableau 36. Valeurs de $F_1$, $F_2$ et $F_0$ pour le PRS (les trois bassins)..... 72

Tableau 37. Valeurs estimées des paramètres du modèle 3.6 pour le PRS (les trois bassins)..... 72

Tableau 38. Résultats de l'ajustement du modèle 3.4 aux mesures de la profondeur au disque de Secchi - lac Erié, 1967 à 1978..... 73

Tableau 39. Valeurs de la statistique F pour le test d'égalité d'une année à l'autre des cycles des profondeurs au disque de Secchi..... 74

Tableau 40. Résultats de l'ajustement du modèle 3.4 aux données sur les mesures de turbidité - lac Erié, 1967 à 1972..... 75

Tableau 41. Valeurs de la statistique F pour le test d'égalité d'une année à l'autre des cycles saisonniers de la turbidité..... 76

Tableau 42. Rapport de l'ammoniac au nitrite et au nitrate.... 76

Tableau 43. Résumé des concentrations d'oxygène dissous et des profondeurs pour les groupes restreints à des stations communes au cours d'une même année, bassin central, 1967 à 1979..... 121

Tableau 44. Résumé des concentrations d'oxygène dissous et des profondeurs pour les groupes restreints à des stations communes au cours d'une même année, bassin est, 1967 à 1980..... 122

Tableau 45. Résumé des paramètres de régression linéaire simple et des indicateurs de précision de l'analyse de régression des concentrations d'oxygène dissous, bassins central et est..... 123

Page

Tableau 46. Estimation des pentes et des ordonnées à l'origine obtenues par régression avec pondération..... 124

Tableau 47. Estimation des pentes et des ordonnées à l'origine obtenues par régression avec poids arbitraires.. 124

Tableau 48. Résultats des tests d'égalité des pentes et des ordonnées à l'origine - Modèle 4.1..... 125

Tableau 49. Estimations des paramètres du modèle 5.2..... 145

Tableau 50. Résumé du nombre de stations où les concentrations de coliformes totaux ou fécaux, ou les deux, ont été déterminées, ceci pour toutes les données de surface disponibles..... 197

Tableau 51. Distributions des fréquences observées, f(r), des concentrations de coliformes totaux, r, lors des missions 111 de 1966, 102 de 1968 et 107 de 1970..... 198

Tableau 52. Précision des tests d'ajustement de la distribution binomiale négative aux données sur les coliformes totaux de 1968, basés sur des estimations par MM et MV de p et k..... 199

Tableau 53. Comparaison des méthodes de regroupement des données sur les coliformes totaux..... 200

Tableau 54a. Comparaison de l'ajustement de la distribution binomiale négative à certaines des données sur les coliformes totaux de 1967 effectuée à l'aide de toutes les données de la mission et d'un sous-ensemble dont les valeurs élevées ont été éliminées..... 201

Tableau 54b. Comparaison de l'ajustement de la distribution binomiale à certaines des données sur les coliformes totaux de 1967 effectuée à l'aide de toutes les données de la mission et d'un sous-ensemble dont les valeurs élevées ont été éliminées..... 201

Tableau 55. Test d'ajustement de la distribution binomiale négative aux données sur les coliformes totaux de 1967, 1968 et 1969..... 202

Tableau 56. Estimations des paramètres de la distribution binomiale négative pour les données sur les coliformes totaux de 1967, 1968 et 1969..... 203

Page

Tableau 57. Test d'ajustement de la distribution lognormale aux données sur les coliformes totaux de la mission 111 de 1966 et de la mission 107 de 1967..... 204

Tableau 58. Zones 1 à 5 des coliformes totaux par mission, de 1966 à 1970..... 205

Tableau 59. Concentrations de coliformes totaux et emplacements des stations de la zone 5..... 206

Tableau 60. Distributions des fréquences observées des données sur les coliformes fécaux, 1967 à 1975.. 207

Tableau 61. Distribution des fréquences conditionnelles $f(R_f/R_t)$ des concentrations de coliformes fécaux, $R_f$, pour des concentrations données de coliformes totaux $R_t$, données de 1967 et 1968..... 208

Tableau 62. Résumé des concentrations de coliformes fécaux aux stations où les concentrations de coliformes totaux et fécaux ont été déterminées..... 209

Tableau 63. Résumé des concentrations de coliformes totaux aux stations où les concentrations de coliformes totaux et fécaux ont été déterminées. 210

Tableau 64. Rapport des concentrations de coliformes fécaux aux concentrations de coliformes totaux..... 211

Tableau 65. Turbidité et coliformes totaux par zones utilisant seulement les stations où la turbidité et les coliformes totaux ont été déterminés..... 212

Tableau 66. Moyenne, moyenne pondérée, variance, efficacité et nombre de stations pour chaque mission..... 248

Tableau 67. Affectation optimale des stations d'échantillonnage du lac Erié..... 248

Tableau 68. Ecart relatif et taille de l'échantillon correspondante permettant de déceler des différences au niveau de signification de 5 %... 249

# Illustrations


| | | Page |
|---|---|---|
| Figure 1. | Les Grands lacs | 24 |
| Figure 2. | Lac Erié, bathymétrie | 25 |
| Figure 3. | Concentration de phosphore dans le lac Erié, 1970 à 1980 | 26 |
| Figure 4. | Taux moyen d'appauvrissement en oxygène dissous durant l'été dans les eaux de fond, partie centrale du lac Erié | 27 |
| Figure 5. | Taux chronologique d'appauvrissement en oxygène dissous dans l'hypolimnion du centre du lac Erié, normalisé selon l'épaisseur en 1970 | 27 |
| Figure 6. | Limites des bassins ouest, central et est du lac Erié | 77 |
| Figure 7. | Valeurs de EC en fonction du mois et de la température (1968-1976) | 77 |
| Figure 8. | Niveau annuel moyen de l'eau du bassin central | 78 |
| Figure 9. | Médiane mobile de cinq ans | 78 |
| Figure 10. | Somme des carrés résiduelle en fonction de l'année | 78 |
| Figure 11. | Niveaux d'eau observés et estimés en fonction de l'année | 79 |
| Figure 12. | Résidus normalisés et orthogonalisés en fonction de l'année | 79 |
| Figure 13. | Résidus normalisés cumulés en fonction de l'année | 80 |
| Figure 14. | Résidus au carré cumulés et leurs valeurs prévues | 80 |
| Figure 15. | Valeurs de $\hat{\alpha}_0$ en fonction de l'année | 81 |
| Figure 16. | Valeurs de $\hat{\alpha}_1$ en fonction de l'année | 81 |
| Figure 17. | Valeurs de $\hat{\alpha}_2$ en fonction de l'année | 82 |

Page

Figure 18. Niveaux d'eau observés et estimés à l'aide du modèle 3.3..... 82

Figure 19. Niveau d'eau moyen au printemps, en été, en automne et en hiver..... 83

Figure 20. Débit annuel moyen de la rivière Niagara à la hauteur de Queenston..... 83

Figure 21. Relation entre le débit de la rivière Niagara et le niveau du lac Erié..... 84

Figure 22a. Moyennes mensuelles des températures de l'air de 1967 à 1978 (bassin central)..... 85

Figure 22b. Courbes des RC des températures de l'air de 1967 à 1978 (bassin central)..... 85

Figure 23. Cycles saisonniers des températures dans les bassins ouest, central et est..... 86

Figure 24. Température moyenne de l'eau de surface du bassin ouest en fonction de celle des bassins central et est pendant les périodes de réchauffement et de refroidissement..... 86

Figure 25. Températures moyennes de l'eau de surface et valeurs estimées à l'aide du modèle 3.4..... 87

Figure 26a. Résidus moyens des températures du bassin ouest..... 87

Figure 26b. Résidus moyens des températures du bassin central... 88

Figure 26c. Résidus moyens des températures du bassin est..... 88

Figure 27. Profil bathythermographique (29 juillet au 3 août 1968)..... 89

Figure 28. Profil bathythermographique du bassin central, 1967. 90

Figure 29. Profil bathythermographique du bassin central, 1978. 91

Figure 30. Températures moyennes de l'épilimnion et de l'hypolimnion du bassin central de 1967 à 1978 et droites de régression ajustées aux données sur l'hypolimnion..... 92

Figure 31. Teneurs moyennes non corrigées en chlorophylle a en fonction du temps..... 92

Page

Figure 32. Estimations de l'effet de l'année en fonction des années (chlorophylle a)..... 93

Figure 33. Teneurs moyennes en PT en fonction du jour de l'année dans les bassins ouest, central et est.... 93

Figure 34. Estimations de l'effet de l'année en fonction des années (PT)..... 94

Figure 35. Effet de l'année sur la chlorophylle a en fonction de celui sur le PT..... 94

Figure 36. Tendances des années sur le PRS..... 94

Figure 37. Teneurs en $N\text{-}NO_3NO_2$ observées et estimées dans le bassin ouest..... 95

Figure 38a. Teneurs en $N\text{-}NO_3NO_2$ observées et estimées dans le bassin ouest en 1967, 1968 et 1970..... 95

Figure 38b. Teneurs en $N\text{-}NO_3NO_2$ observées et estimées dans le bassin ouest en 1971..... 96

Figure 38c. Teneurs en $N\text{-}NO_3NO_2$ observées et estimées dans le bassin ouest en 1978..... 96

Figure 39a. Teneurs en $N\text{-}NO_3NO_2$ observées et estimées dans le bassin central en 1967, 1968 et 1970..... 97

Figure 39b. Teneurs en $N\text{-}NO_3NO_2$ observées et estimées dans le bassin central en 1971..... 97

Figure 39c. Teneurs en $N\text{-}NO_3NO_2$ observées et estimées dans le bassin central en 1978..... 98

Figure 40. Teneurs en $N\text{-}NO_3NO_2$ observées et estimées dans le bassin est..... 98

Figure 41. Teneurs en $NH_3$ observées et estimées dans le bassin ouest..... 99

Figure 42. Teneurs en $NH_3$ observées et estimées dans le bassin central..... 99

Figure 43. Teneurs en $NH_3$ observées et estimées dans le bassin est..... 100

Figure 44. Mesure de la profondeur estimée au disque de Secchi dans chacun des bassins en fonction de l'année..... 101

Page

Figure 45. Profondeurs observées au disque de Secchi et profondeurs estimées à l'aide du modèle 3.4..... 101

Figure 46. Résidus moyens pour les profondeurs au disque de Secchi..... 102

Figure 47. Relation entre l'effet de l'année sur la chlorophylle a et l'effet de l'année sur le PT.... 102

Figure 48. Organigramme de l'algorithme de regroupement en grappes..... 126

Figure 49. Exemple de tracés générés par le programme de regroupement, mission 7022106..... 127

Figure 50. Ecart-type de la concentration d'oxygène dans l'hypolimnion pour les grappes du bassin est en fonction du jour du calendrier..... 127

Figure 51. Reports des résidus dans l'ordre des observations au bassin central: (a) résidus normalisés et orthogonalisés; (b) résidus cumulés; (c) résidus orthogonalisés au carré..... 128

Figure 52. Reports des résidus dans l'ordre des observations au bassin est: (a) résidus normalisés et orthogonalisés; (b) résidus cumulés; (c) résidus orthogonalisés au carré..... 129

Figure 53. Taux d'appauvrissement en oxygène dissous en fonction de l'année et courbes ajustées, bassins central et est..... 130

Figure 54. Comparaison des taux d'appauvrissement en oxygène dissous non corrigés tels que calculés par divers auteurs..... 130

Figure 55. Ordonnée à l'origine des droites de régression de la concentration d'oxygène dissous en fonction du niveau du lac Erié..... 146

Figure 56. Taux d'appauvrissement estimé en fonction du niveau du lac Erié..... 146

Figure 57. Taux d'appauvrissement estimé selon le modèle 5.2 en fonction du niveau du lac Erié..... 146

Figure 58. Tracé des concentrations d'oxygène observées et de leurs valeurs estimées selon le modèle 5.10 en fonction du temps en jours du calendrier..... 148

Page

Figure 59. Tracé du taux d'appauvrissement en fonction du niveau de l'eau pour diverses teneurs en PT..... 147

Figure 60. Intervalle de confiance à 95 % pour les teneurs en PT, de 1948 à 1962..... 148

Figure 61 Concentration de $0_2$ prévue par le modèle 5.10 en fonction des jours du calendrier pour plusieurs températures, niveaux de l'eau et teneurs en PT... 149

Figure 62. Probabilité d'anoxie pour une période de stratification de 110 jours et diverses teneurs en PT et niveaux d'eau..... 149

Figure 63. Distributions des fréquences observées et distributions de Poisson et binomiale négative ajustées, données sur les coliformes totaux de 1968..... 213

Figure 64. Isoligne de vraisemblance relative R(p,k) = c, c = 0.01, 0.05, 0.50, 0.90, 0.99, pour les paramètres p et k de la distribution binomiale négative et les données sur les coliformes totaux - missions 102, 108, 109 et 112 de 1968.... 214

Figure 65a. Fonctions de maximum de vraisemblance relatif de $R_m(p)$ en fonction de p, un paramètre de la distribution binomiale négative, et la concentration de coliformes totaux - missions 102, 108, 109 et 112 de 1968..... 215

Figure 65b. Fonctions de maximum de vraisemblance relatif de $R_m(k)$ en fonction de k, un paramètre de la distribution binomiale négative, et la concentration de coliformes totaux - missions 102, 108, 109 et 112 de 1968..... 215

Figure 66. Distributions des probabilités correspondant aux groupes déterminés pour la mission 102 (1968)..... 216

Figure 67. Distributions des fréquences observées des concentrations de coliformes totaux (R) ou de $\log_e R$ présentées selon deux échelles..... 217

Figure 68. Concentrations moyennes de coliformes totaux des zones 1 à 4 et des zones 1 et 2, et estimations par maximum de vraisemblance des paramètres de la distribution binomiale négative pour chaque mission..... 219

Page

Figure 69. Zones du lac Erié, selon les concentrations de coliformes totaux, par mission, de 1966 à 1970.... 220

Figure 70. Exemples de zones 2 caractéristiques pour les distributions de coliformes totaux de printemps et maximales..... 224

Figure 71. Zones 3, 4 et 5 selon les concentrations de coliformes totaux regroupées pour toutes les missions de 1967..... 224

Figure 72. Concentrations de coliformes fécaux en fonction de deux gammes de concentrations de coliformes totaux - missions de 1967 et de 1968..... 225

Figure 73. Distributions des fréquences conditionnelles des concentrations de coliformes fécaux pour une concentration donnée de coliformes totaux ($R_t$) et $R_t \leq 20$, lors de trois missions de 1967 et de trois missions de 1968..... 226

Figure 74. Concentrations de coliformes fécaux en fonction des concentrations de coliformes totaux par zone et par mission pour 1967 et 1968..... 227

Figure 75. Concentration moyenne de coliformes fécaux en fonction de la concentration moyenne de coliformes totaux dans la zone 1 pour les missions de 1967 à 1969..... 228

Figure 76. Rapports de la concentration des coliformes fécaux à celle des coliformes totaux en fonction de la date des missions - zones 1, 2 et 3, données des stations communes..... 228

Figure 77. Emplacements des stations où les concentrations de coliformes fécaux ont été déterminées..... 229

Figure 78. Fonction de vraisemblance conditionnelle relative $R_c(\rho)$ pour le rapport des concentrations de coliformes fécaux et totaux lors de quatre missions..... 231

Figure 79. Concentrations de coliformes fécaux dans les zones de coliformes totaux - deux exemples..... 231

Figure 80. Concentration de coliformes totaux en fonction de la turbidité dans le bassin ouest pendant les missions de 1968..... 232

Page

Figure 81. Concentration de coliformes totaux en fonction de la turbidité lors des missions de 1968..... 232

Figure 82. Concentration moyenne de coliformes totaux en fonction de la turbidité moyenne lors des missions de 1968 aux stations communes des zones 1 à 4 déterminées à partir de l'ensemble des données sur les coliformes totaux..... 233

Figure 83. Nombre de stations nécessaire pour déceler une différence de la moyenne à un niveau de signification de 5 % en fonction des déviations relatives..... 250

Figure 84. Variance de la température moyenne de surface en fonction du nombre de stations..... 251

Figure 85. Variance de la chlorophylle a en fonction du nombre de stations..... 252

Figure 86. Déviation en pourcentage pour la chlorophylle a moyenne pour divers formats d'échantillons..... 252

# Résumé

Le présent rapport est une évaluation statistique des données relatives à la qualité de l'eau du lac Erié obtenues de 1966 à 1981 dans le cadre du programme de surveillance des Grands lacs. On trouvera un résumé au premier chapitre et, au chapitre 2, une mise à jour du compte rendu des premiers travaux portant sur l'eutrophisation du lac Erié. Le troisième chapitre traite du problème de l'examen des données pour déterminer l'existence et la nature de toute tendance que pourraient présenter les valeurs des paramètres limnologiques suivants: (1) le niveau de l'eau du lac Erié entre 1900 et 1979; (2) le débit de la rivière Niagara entre 1860 et 1975; et (3) la température de l'air et de l'eau, la profondeur au disque de Secchi, la turbidité et les teneurs en phosphore total, en phosphore réactif soluble, en chlorophylle a et en azote pendant la période allant de 1967 à 1980. Une analyse est effectuée pour chacun des bassins, ouest, central et est, du lac Erié. On présente au chapitre 4 une technique d'analyse par grappes qui permet de regrouper automatiquement les données selon les divers régimes thermiques et spatiaux naturels. Il s'avère que les taux d'utilisation de l'oxygène dissous calculés à partir des stations visitées à toutes les sorties dans ces grappes pendant une année sont compris dans ceux calculés par les chercheurs précédents. L'analyse par grappes est donc une solution de remplacement semi-objective et pratique pour les méthodes antérieures de choix de données. Divers types d'analyse par régression ont été utilisés pour déterminer les taux d'épuisement de l'oxygène hypolimnétique au cours des années 1967 à 1980 dans les bassins du centre et de l'est du lac Erié. Des tests statistiques ont montré que le taux d'épuisement a varié d'une année à l'autre. On traite, au chapitre 5, du problème que pose

l'explication de ces différences annuelles. Le modèle utilisé fait appel au niveau de l'eau, à la température de l'hypolimnion et aux concentrations moyennes de phosphore total annuelles comme variables explicatives. Le modèle permet d'estimer la probabilité d'anoxie en fonction de ces trois variables. On traite aussi de l'utilisation du modèle pour l'élaboration de réglementations et la détermination de normes. Les données obtenues entre 1966 et 1970 ont servi à caractériser les variabilités spatiales et temporelles des teneurs de coliformes totaux et fécaux du lac Erié. De plus, on s'intéresse à la relation entre la turbidité et la teneur de coliformes totaux. On présente aussi, au même chapitre, certaines méthodes statistiques générales utiles à l'analyse de données discrètes (dénombrements), y compris d'autres variables limnologiques comme le phytoplancton, le zooplancton et le poisson. On s'intéresse, au chapitre 7, à la façon d'utiliser les vieilles informations sur la variabilité spatiale et temporelle des données limnologiques dans le but de planifier une stratégie pour la cueillette des données quand on désire estimer la valeur moyenne, pondérée en fonction de la superficie, d'une seule variable limnologique. Cette approche statistique est appliquée aux dénombrements des coliformes, à la température et aux teneurs de chlorophylle a.

# Abstract

This report presents a statistical evaluation of the water quality data available on Lake Erie from the Great Lakes surveillance program, 1966 to 1981. In Chapter 1, an introduction is given. Chapter 2 presents an up-to-date review of the early historical work on the eutrophication of Lake Erie. Chapter 3 discusses the problem of examining the data to determine the existence and nature of any trend that may be present in the values of the following limnological parameters: (1) the water level of Lake Erie for the years 1900 to 1979; (2) Niagara River flows for the period 1860 to 1975; (3) air and water temperature, Secchi disk depth, turbidity, total phosphorus, soluble reactive phosphorus, chlorophyll a, and nitrogen for the period 1967 to 1980. The analysis is performed for each of the Western, Central and Eastern basins of Lake Erie. Chapter 4 presents a clustering technique which automatically separates the data into the various naturally occurring thermal and spatial regimes. The resulting dissolved oxygen depletion rates, which are calculated on the basis of common stations in these clusters within a year, are found to be between those calculated by previous authors. Thus the cluster analysis presents a semi-objective and practicable alternative to the previous methods of data selection. Various types of regression analysis were applied to determine the hypolimnetic oxygen depletion rates for the years 1967 to 1980 in the Central and Eastern basins of Lake Erie. Statistical tests indicated that the depletion rate changed from year to year. The problem of explaining the differences in the depletion rate from year to year is considered in Chapter 5. The model uses the lake water level, the hypolimnion temperature and yearly mean total phosphorus concentrations as explanatory variables. It enables the estimation of the probability

of anoxia as a function of the three explanatory variables. Also, the use of the model for setting regulations and for specifying standards is given. The characterization of the spatial and temporal variabilities of total coliform and fecal coliform concentrations in Lake Erie is determined using the available data for 1966 to 1970. Furthermore, the relationship between turbidity and total coliform concentration is considered. Also in this chapter, some general statistical methodologies are given which are useful for the analysis of discrete data (i.e. counts) including other limnological variables such as phytoplankton, zooplankton and fish. In Chapter 7, the question is investigated concerning how to use past information about the spatial and temporal variabilities of limnological data to plan a strategy for future data collection when the aim is to estimate the areal weighted mean value of a single limnological variable. This statistical approach is applied to coliform counts, temperature and chlorophyll a.

# Remerciements

Les auteurs du présent rapport tiennent à remercier tous ceux qui, directement ou indirectement, ont collaboré à sa préparation. Nous tenons à remercier tout particulièrement K. Rodgers, F. Elder, J. Barica, R.A. Vollenweider, B.J. Dutka, P. Hamblin, C.R. Murthy, D.C. Lam et D. Warry de leurs encouragements, de leur aide et de leurs conseils.

Nous remercions M. Fellows et A. Liu de l'aide apportée pour la programmation. Nous désirons, pour terminer, remercier C.L. Minnie pour son excellente dactylographie du texte, et W.D. Finn et ses collaborateurs pour la grande qualité des figures.

Nous voulons aussi exprimer notre reconnaissance pour l'aide financière obtenue dans le cadre de l'Accord sur la qualité de l'eau dans les Grands lacs.

# Introduction

Le présent rapport a pour objectif de résumer, à l'aide de techniques statistiques, les données sur le lac Erié recueillies de 1966 à 1981 dans le cadre du Programme de surveillance des Grands lacs. Ce programme a été conçu et mené par le gouvernement canadien afin d'obtenir de l'information détaillée et semi-systématique sur les modifications spatiales et temporelles survenues dans la qualité de l'eau des Grands lacs. On ne disposait, avant l'existence du programme, que d'informations clairsemées sur les processus biologiques, chimiques et physiques se déroulant dans les lacs. Les données obtenues ont certainement contribué à accroître nos connaissances de la dynamique de la plupart des caractéristiques limnologiques et de leurs interactions. De plus, ces données nous ont permis de reconsidérer et de modifier nos concepts et de tester d'autres modèles et hypothèses relatives aux facteurs qui régissent la qualité de l'eau. Seule une petite partie de l'information contenue dans les données de surveillance a fait l'objet d'une étude scientifique. Cela s'explique par le fait que la plupart des études ont été restreintes aux caractéristiques d'un petit nombre de variables limnologiques, n'ont porté que sur les données provenant d'un certain nombre de missions ou faisaient appel à des méthodes inadéquates pour l'analyse des données. Une évaluation statistique complète et intégrée de toutes les données obtenues, par opposition à des études limitées et isolées, assurera l'utilisation de toute l'information générée par le programme de surveillance. Une telle analyse permettra d'isoler les diverses sources de variabilité (spatiale et temporelle) et d'étudier les relations entre les différentes variables grâce à des modèles empiriques. De plus, le programme de surveillance a été lancé il y a plus de quinze ans et il

s'avère maintenant opportun d'en considérer les possibilités et d'évaluer l'utilité de l'information. Cela permettra d'apporter en temps et lieu les modifications nécessaires au programme.

De façon plus spécifique, le présent rapport traite:

(1) de l'extraction de l'information contenue dans les données de surveillance et de son utilisation pour (a) l'examen des variations de la qualité de l'eau (c'est-à-dire la détermination de la tendance temporelle), (b) l'étude de la variabilité spatiale et la délimination de régions des Grands lacs présentant une eau de faible qualité et (c) l'élaboration de modèles empiriques et l'étude de la relation entre les différentes variables limnologiques;

(2) de l'élaboration d'une stratégie pour la cueillette des données;

(3) de l'élaboration de méthodes pouvant être appliquées à des problèmes écologiques similaires, comme l'étude de la variabilité spatiale et temporelle des paramètres mesurés afin d'étudier les effets des précipitations acides.

# Rétrospective du lac Érié

*par R.E. Kwiatkowski*

Le Conseil international sur la pollution des eaux du lac Erié et le Conseil sur la pollution des eaux du lac Ontario et du fleuve Saint-Laurent (1969) ont recommandé que les gouvernements du Canada et des Etats-Unis élaborent en collaboration un programme de lutte contre la pollution de l'écosystème des Grands lacs. Trois éléments (phosphore, azote et silicium) se sont avérés des facteurs limitatifs de la productivité des plantes aquatiques dans les Grands lacs. La Commission mixte internationale (CMI) s'est fortement intéressée aux moyens de lutter contre la pollution par le phosphore, cette matière nutritive étant la plus facile, tant techniquement qu'économiquement, à éliminer des plans d'eau. Le Canada et les Etats-Unis ont donc signé, en 1972, l'Accord sur la qualité de l'eau dans les Grands lacs (AQEGL, 1972). On y reconnaissait que l'eutrophisation posait un problème sérieux dans les lacs d'aval (Erié et Ontario) et la CMI était chargée d'élaborer des objectifs relatifs aux charges de phosphore afin de faciliter la tâche au Comité sur la qualité de l'eau des Grands lacs. Un examen exhaustif des programmes en vigueur a été effectué cinq ans après la signature de l'Accord et un nouvel accord où des objectifs relatifs aux charges de phosphore total pour chacun des Grands lacs étaient précisés a été signé en 1978 (AQEGL, 1978). Contrairement à l'approche utilisée dans l'accord de 1972, les objectifs de l'accord de 1978 reposent sur la qualité de l'eau correspondant aux charges de phosphore dans chacun des bassins.

La remise en état aérobie, tout au long de l'année, des eaux du fond du bassin central du lac Erié a été énoncé comme objectif premier à l'annexe 2 de l'Accord sur la qualité des eaux de 1972 et à l'annexe 3 de l'Accord de 1978. D'importantes recherches scientifiques ont été effectuées au cours des dix dernières années afin d'atteindre ce but. Nous faisons ici un bref examen des premiers travaux portant sur les gaz dissous dans les eaux des lacs et sur les facteurs régissant l'épuisement de l'oxygène dans les eaux hypolimnétiques et nous passons en revue certains des principaux travaux portant sur l'étendue de l'anoxie dans le lac Erié. Il n'est pas dans notre intention de faire ici un examen détaillé ou complet de l'épuisement de l'oxygène dans les lacs tempérés, cela débordant du cadre de notre étude. Nous désirons cependant permettre au lecteur de mieux apprécier les interactions biologiques, chimiques et physiques complexes qui prennent place dans les lacs ainsi que la quantité importante d'information obtenue sur le problème de l'eutrophisation liée à l'épuisement de l'oxygène dans les eaux hypolimnétiques du lac Erié, le plus petit des Grands lacs (tableau 1, figure 1).

Le lac Erié est, en réalité, formé de trois lacs. La plus grande partie du lac est peu profonde; la profondeur du petit bassin ouest est généralement inférieure à 11 m et la profondeur maximale du grand bassin central est d'environ 25 m. Le bassin est s'avère le plus profond du lac, la profondeur pouvant atteindre 60 m (figure 2). On pourra trouver dans Sly (1976) le détail de la géologie du socle rocheux des trois bassins et dans Simons (1976) de l'information relative au transport de l'eau entre les bassins.

## PREMIERS TRAVAUX SUR LES GAZ DISSOUS

Hoppe-Seyler (1895) a été le premier à traiter de l'importance de la perte de $O_2$ des eaux hypolimnétiques en tant que mesure de l'activité biologique. A partir de mesures des

concentrations de $O_2$ à seulement cinq profondeurs (maximum 245 m) du lac Constance, il a réalisé l'importance de l'étude de l'oxygène pour comprendre les processus biologiques se déroulant dans l'eau; c'est ainsi que le concept du déficit d'oxygène est apparu en limnologie. En 1911, Birge et Juday ont traité du cycle annuel de la stratification thermique ainsi que de l'utilisation et de la production de $O_2$ et de $CO_2$ dans les processus biologiques au sein des lacs. Birge (1906) avait déjà écrit que les pertes de $O_2$ de l'hypolimnion étaient fonction de quatre principaux facteurs: la quantité de matières décomposables (autochtones et allochtones) en provenance de l'épilimnion; le volume de l'hypolimnion (lui-même fonction de la profondeur du lac); la durée pendant laquelle les eaux du fond sont séparées de l'épilimnion et la température des eaux du fond (qui a un effet sur la vitesse de décomposition).

August Thienemann a utilisé les valeurs du déficit d'oxygène pour la mesure quantitative du degré d'eutrophisation (eutrophe: riche en matières nutritives; oligotrophe: pauvre en matières nutritives) d'un lac. Thienemann (1915) a déterminé cinq facteurs régissant l'importance de la perte d'oxygène de l'hypolimnion: (1) la saison; (2) la situation du lac (en fonction des vents dominants); (3) le volume de l'hypolimnion et le rapport entre les volumes d'eau se trouvant au-dessus et au-dessous de la thermocline; (4) la température des eaux hypolimnétiques et (5) la quantité de matière organique transportée vers les eaux du fond. Thienemann (1928) a, par la suite, indiqué que le troisième facteur était vraiment important car il donnait une signification particulière au rapport du volume de l'épilimnion à celui de l'hypolimnion. De plus, il a conclu qu'étant donné que les eaux du fond des lacs peu profonds se réchauffaient, toutes choses égales d'ailleurs, plus rapidement que celles des lacs profonds et que la chaleur favorisait la décomposition, la différence de réchauffement due aux différentes profondeurs exerçait une influence similaire à celle due à la

différence des volumes des couches tropholytiques.* Thienemann (1928) a déterminé une caractéristique morphologique pour les lacs oligotrophes: le volume de l'hypolimnion est important par rapport à celui de l'épilimnion (tableau 2), tandis que dans un lac eutrophe, le volume de l'hypolimnion est relativement faible par rapport à celui de l'épilimnion.

## VARIABLES PHYSIQUES, BIOLOGIQUES ET CHIMIQUES REGISSANT LA PERTE DE L'OXYGENE DISSOUS

La thermocline s'oppose au déplacement des gaz dissous de l'épilimnion vers l'hypolimnion. Le cycle annuel de température s'avère donc important dans l'étude de la perte d'oxygène. Les lacs des régions tempérées se couvrent de glace en hiver. Il peut s'établir une stratification permanente (stagnation d'hiver); la température de l'eau immédiatement sous la glace est de 0°C, tandis que celle des strates les plus profondes est uniforme et dépasse à peine 4°C. Après la débâcle, les courants dus au vent mélangent toute la masse d'eau (inversion de printemps) et le lac devient isotherme et chimiquement homogène. A partir de l'équinoxe du printemps, les couches supérieures du lac absorbent de l'énergie thermique. Il a cependant été montré que l'absorption de rayonnement solaire plus intense au printemps ne représentait qu'une petite partie de la chaleur totale dissipée dans le lac (Birge et Juday, 1921). La principale source d'énergie est le vent (Birge, 1916) qui donne naissance à des courants. La vitesse et l'orientation de ces derniers sont fonction de la force et de la direction du vent (Hutchinson, 1967). Quand les particules d'eau de la surface, poussées par le vent, atteignent la rive, elles sont déviées par les

---

*Trophogène - couche superficielle où la production organique est basée sur l'énergie solaire; zone tropholytique - couche profonde où la désassimilation prédomine à cause d'une insuffisance de lumière.

eaux profondes plus froides et plus denses. Il s'établit alors, immédiatement sous la surface, un contre-courant qui provoque un échange thermique (diffusion par turbulence). L'accumulation progressive de chaleur donne naissance à des gradients de densité (l'eau est la plus dense à 4°C) et il y a apparition d'une couche limite (thermocline), dans les lacs suffisamment prodonds, entre la couche de surface à mélange total (épilimnion) et les masses d'eau calme (hypolimnion) sous-jacentes (Birge, 1910). Birge (1898) a défini la thermocline comme la région de baisse de température rapide, soit plus de 1°C par mètre. Brönsted et Wesenberg-Lund (1911) ont redéfini ce terme comme signifiant le plan du taux maximum de diminution de la température, tandis que Hutchinson (1957) désigne par métalimnion toute la région de fort gradient de température, allant du plan supérieur de courbure maximale, nommé le «genou» de la thermocline, au plan inférieur de courbure (inverse) maximale. Il faut souligner qu'étant donné les interactions complexes existant entre le rayonnement solaire, le vent et la morphologie d'un lac, chaque lac peut présenter une courbe de température qui lui est propre et que, les conditions météorologiques variant d'une année à l'autre, un lac donné peut présenter une courbe de température annuelle différente (tableau 3), et, donc, un épilimnion d'épaisseur variable.

Il n'y a plus de stratification thermique dans un lac quand la quantité de chaleur perdue est plus importante que la quantité de chaleur absorbée. Les pertes thermiques par rayonnement sont les plus importantes, mais il y a aussi des pertes par évaporation et conduction aux dépens de l'air et des sédiments, ce qui peut se traduire par des courants de convection verticaux. Le volume d'eau s'écoulant au travers du lac, et qui emporte généralement les eaux les plus superficielles, peut aussi avoir un effet appréciable sur le bilan thermique (Ruttner, 1952). Vers la fin de l'automne, il y a retour des conditions isothermes et le lac redevient homogène (inversion d'automne).

L'importance que revêt la thermocline pour la biologie d'un lac a tout d'abord été décrite dans une communication qui est maintenant devenue un classique: «The Thermocline and Its Biological Significance» (Birge, 1904). Le processus fondamental de la vie, l'assimilation du carbone, se déroule dans les eaux de la partie supérieure du lac (zone trophogène). La vitesse et l'importance de l'assimilation sont régies par une série de facteurs complexes qui dépendent des espèces phytoplanctoniques présentes, du rayonnement solaire, de la disponibilité des matières nutritives, de la température et des relations entre les besoins des organismes du phytoplancton. Ces interactions complexes ont été passées en revue par Hutchinson (1967) et Lund (1965), et nous n'en traiterons pas. La matière organique formée finit par précipiter hors de la zone trophogène, dont l'épaisseur est souvent équivalente à celle de l'épilimnion (Thienemann, 1982), et atteint la zone tropholytique profonde. Ce déplacement de la matière organique se traduit par une diminution de la teneur en oxygène dissous des eaux hypolimnétiques due à la respiration des végétaux, des animaux et des bactéries (décomposition de la matière organique) ainsi qu'à une oxydation purement chimique de la matière organique en solution. La thermocline interdisant le déplacement de l'oxygène dissous de l'épilimnion, il y a donc apparition d'un déficit d'oxygène dans l'hypolimnion. Il n'y a pas toujours oxydation totale de la matière organique dans l'hypolimnion (demande d'oxygène). D'importantes quantités de matières organiques des lacs eutrophes se déposent sur les fonds où elles sont oxydées (demande d'oxygène des sédiments) ou emmagasinées. Il y a apparition d'une microstratification de l'eau désoxygénée entre les sédiments du fond et les eaux susjacentes. Cette microcouche croit graduellement par la diffusion de turbulence et, dans les lacs fortement eutrophes, tout l'hypolimnion peut devenir anoxique. L'anoxie de l'hypolimnion n'a pas que des effets nocifs pour la biologie de cette zone (benthos, zooplancton, bactéries et poissons d'eau froide faisant l'objet d'une pêche), elle en modifie aussi la chimie. Une réduction appréciable de la teneur

en oxygène dissous à l'interface vase-eau se traduit par le passage en solution d'hydroxydes ferriques et manganiques. La réduction du phosphate ferrique se traduit par le retour du phosphore des sédiments vers la masse d'eau (charge interne), ce qui contribue à eutrophiser encore plus le lac.

## ETAT ACTUEL DU LAC ERIE

Une étude limnologique du lac Erié a été effectuée en 1928 et 1929 suite à un déclin sérieux de certaines populations de poissons (Fish, 1960). De mai à septembre, 62 stations ont fait l'objet de prélèvements, lors de huit missions portant sur divers paramètres physiques, chimiques et biologiques. Cette étude a montré que le pourcentage de saturation d'oxygène de l'hypolimnion était de 83.3 % en été. Burkholder (1960) a conclu que le maintien de la concentration d'oxygène dans le lac Erié, en 1929, s'expliquait probablement par le caractère oligotrophe général du lac. Le degré de décomposition s'avérait modéré de sorte que l'on n'a pas noté, comme c'est le cas dans les lacs eutrophes, de faibles teneurs en oxygène. Vers la fin des années soixante, l'anoxie de l'hypolimnion du lac Erié s'était généralisée et il était courant de qualifier ce lac de lac mort.

Des espèces de valeur commerciale, comme le corégone, le cisco de lac et le doré noir, étaient devenues rares ou avaient fait place à des espèces moins intéressantes du point de vue commercial comme le gaspareau, l'éperlan et le malachigan (Baldwin et Saalfield, 1962; Leach et Nepsy, 1976). Leach et Nepsy (1976) ont conclu, après étude de la littérature, que ces changements d'espèces étaient dus à diverses contraintes dont ils ont fait la liste chronologique (et par importance) suivante: pêche commerciale intensive; charge de matières nutritives (provoquant l'anoxie de l'hypolimnion); introduction d'espèces étrangères; aménagement de tributaires et de rives; érosion et envasement; introduction de matières toxiques.

Beeton (1965) a conclu, dans une étude de l'eutrophisation des Grands lacs, que les activités humaines ont nettement accéléré le taux d'eutrophisation du lac Erié. Ses conclusions sont basées sur une analyse des tendances de données chimiques portant sur la période de 1854 à 1960. Beeton (1965) a noté d'importantes augmentations des concentrations de chlorure et de sulfate, substances manifestement présentes dans les déchets domestiques et industriels. La concentration de ces paramètres a varié parallèlement à la croissance démographique dans le bassin du lac Erié. La comparaison de données récentes pour la zone riveraine du bassin central avec celles de Beeton (1965) a montré que les concentrations notées pour 1978-1979 étaient semblables à celles notées vers la fin des années cinquante et que les teneurs en chlorure et en calcium étaient même plus faibles (Richards, 1981). Richards a conclu que le lac ne se détériorait pas au même rythme qu'au cours de la première moitié du siècle.

Une étude biologique de la communauté benthique du bassin ouest du lac Erié effectuée par Carr et Hiltunen (1965) a montré que les conditions environnementales étaient plus difficiles en 1961 qu'elles ne l'étaient en 1930. La population de l'éphémère fouisseur Hexagenia spp. est passée d'une moyenne de 139/$m^2$ en 1930 à moins de 1/$m^2$ en 1961.* Les oligochètes, un taxon type des milieux peu oxygénés, ont vu leur population s'accroître par un facteur de neuf au cours de la même période (de 677/$m^2$ en 1930 à 5949/$m^2$ en 1961; Carr et Hiltunen, 1965). Mais comme l'a souligné le Groupe de

---

*La CMI signale, dans son rapport de 1980 sur la qualité des eaux des Grands lacs, que les études benthiques du bassin ouest, effectuées en 1979, ont révélé une amélioration appréciable de la communauté benthique. Hexagenia limbata est réapparue pour la première fois depuis le début des années cinquante, près de l'embouchure de la rivière Détroit (CMI, 1980a).

travail sur les stratégies de gestion du phosphore (1980), ces importantes modifications de la communauté benthique se sont produites au cours de périodes pendant lesquelles des substances toxiques, comme le DDT, ont été utilisées pour la première fois, sur une grande échelle, de sorte qu'il est difficile de déterminer des relations de cause à effet.

Les seules études à long terme sur les communautés phytoplantoniques du lac Erié ont été effectuées par Davies (1964, 1969). On a noté, entre les années 1929 et 1962, une augmentation annuelle moyenne de la concentration du phytoplancton, soit une augmentation de 443 cellules/mL/an, due notamment à la prolifération des algues bleues-vertes, groupe généralement associé à l'eutrophisation. Il faut cependant noter que ces données proviennent d'une usine de traitement des eaux située à Cleveland; les modifications de la composition phytoplanctonique signalées par Davies (1964) ne portent donc pas sur les eaux du large, mais sont plutôt représentatives des effets d'apports localisés. Une étude effectuée, en 1981, par Harris et Vollenweider sur les frustules de diatomées présentes dans des carottes prélevées dans le bassin central du lac Erié a montré que la composition en espèces de ce groupe avait, suite au déboisement, beaucoup changé vers 1850. Une deuxième modification mineure, de l'abondance relative et non des espèces, s'est produite au milieu du présent siècle, probablement suite à des rejets accrus de phosphore, sous forme de détersifs, dans le lac Erié. Harris et Vollenweider (1982) ont indiqué que la composition des diatomées était demeurée relativement constante au cours des 100 dernières années.

Des teneurs en phosphore excessives ont été identifiées comme étant le plus important des facteurs d'eutrophisation des Grands lacs qui sont modifiables (Groupe de travail sur les stratégies de gestion du phosphore, 1980). De nombreux auteurs ont, au cours des 20 dernières années, établi des relations

phosphore-chlorophylle-phytoplancton dans les lacs (voir Nichols et Dillon, 1978). Ces relations indiquent qu'une augmentation des charges de phosphore du lac Erié, due aux rejets d'eaux résiduaires et de détersifs et aux apports du ruissellement agricole ou de source atmosphérique, se traduira par une augmentation de la teneur en phosphore des eaux du lac. Cela se traduira à son tour par une augmentation de la biomasse des algues et, par voie de conséquence, par un accroissement de la superficie d'hypolimnion anoxique pendant la stratification d'été.

La CMI a indiqué, en 1977, que les concentrations annuelles moyennes de phosphore s'étaient accrues de façon significative ($P \leq 0.05$) dans le bassin central depuis 1973 (CMI, 1977). Elle a conclu cependant en 1979, après son contrôle annuel des eaux du lac Erié, que les concentrations de phosphore total n'avaient fluctué que légèrement depuis 1970 (CMI, 1979b) et, en 1980, que la concentration de phosphore (figure 3) avait peu varié au cours de la dernière décennie (CMI, 1980b). La mobilité du phosphore, la grande variation de sa concentration d'une année à l'autre, son recyclage et son importante interaction avec les populations biologiques rendent très difficile le calcul des variations de concentration de cette substance dans les Grands lacs (Groupe de travail sur les stratégies de gestion du phosphore, 1980).

La CMI a, depuis 1973, calculé les charges de phosphore total dans les Grands lacs en contrôlant les apports des tributaires, les rejets directs des sources urbaines et industrielles, les canaux de communication et les sources atmosphériques. On a, depuis 1976, fait des estimations pour les sources diffuses. Nous ne disposons malheureusement pas de modes de cueillette coordonnée pour les données relatives aux charges de phosphore, ni de méthodologie normalisée et scientifiquement fiable pour le calcul de ces charges (Groupe de travail sur les stratégies d'élimination du phosphore, 1980). Cela a été démontré par l'exemple du Groupe de travail sur

les diverses charges de phosphore total calculées pour le lac Erié en 1976. Le Groupe de travail III (groupe d'étude canado-américain mis sur pied lors de l'élaboration de l'accord de 1978 sur la qualité de l'eau dans les Grands lacs) a calculé une charge de 19 500 tonnes par année en se basant sur l'étude de la gestion des eaux usées rejetées dans les Grands lacs effectuée par le Army Corps of Engineers des Etats-Unis. Le Comité sur la qualité de l'eau a estimé, à partir de sa série de données, une charge de 15 500 tonnes par an, tandis que le PLUARG calculait une charge de 17 450 tonnes par an. Même si les charges sont calculées par un seul organisme, cela n'assure pas non plus la normalisation des résultats. Le Comité sur la qualité de l'eau a, de 1972 à 1977, utilisé diverses méthodes pour le calcul des charges de phosphore (Zar, 1980). Zar (1980) a conclu à la nécessité de travaux supplémentaires pour évaluer la qualité et l'exactitude statistique des données. Aucune des méthodes d'estimation des charges utilisées pour le calcul de la charge de phosphore total du lac Erié n'a pris en considération la régénération du phosphore des sédiments déposés, même si la quantité ainsi produite peut être équivalente à 111 % de l'apport externe en conditions anoxiques (Burns et Ross, 1972a). La régénération du phosphore est aussi importante en conditions d'oxygénation. On a noté une régénération égale à 137 % de l'apport externe, ceci pour une période de deux mois d'été au lac Erié et dans des conditions d'oxygénation et d'anoxie. Il y a aussi désaccord quant à savoir si les programmes de lutte devraient être basés sur le phosphore total ou le phosphore disponible pour les algues pénétrant dans le plan d'eau (Lee et coll., 1980). Le phosphore se présentant sous un grand nombre de formes de disponibilité très variable, il est pratiquement impossible d'élaborer des modes d'analyse permettant de quantifier chaque forme biologiquement disponible. Les programmes actuels de lutte contre l'eutrophisation sont donc basés sur le concept de la limitation de toutes les formes de phosphore, que celles-ci puissent ou non permettre la croissance des algues (Lee et coll., 1980).

Les craintes de plus en plus importantes liées à l'accroissement de la superficie d'eau anoxique du bassin central du lac Erié ont été à l'origine de la mise sur pied, en 1970, d'une étude canado-américaine intensive (Projet Hypo). Des échantillons d'eau ont été prélevés à 25 stations lors de 10 relevés portant sur le bassin central, et 16 stations supplémentaires ont été établies pour obtenir des enregistrements bathythermiques. Cinq des 25 stations ont été considérées comme principales et ont fait l'objet de prélèvements intensifs portant sur les quatre disciplines (chimie, biologie, bactériologie et physique) considérées comme les plus importantes pour l'étude de la perte d'oxygène (Burns et Ross, 1972a). Burns et Ross (1972b) ainsi que l'étude de 1970 ont conclu qu'il fallait réduire immédiatement des apports de phosphore au lac Erié de façon à améliorer rapidement l'état du lac. On a aussi conclu que, dans le cas contraire, le lac se détériorerait beaucoup plus rapidement qu'au cours des dernières années. L'anoxie de l'hypolimnion du bassin central ne permettait pas, en 1970, la vie des poissons et provoquait une régénération de quantités importantes de phosphore à partir des sédiments (Burns, 1976b).

Beeton (1965) a noté, en 1959 et 1960, de faibles concentrations d'oxygène dissous, mais a mentionné que des observations éparses de faibles teneurs en oxygène avaient été signalées dans le lac Erié au cours des 33 années précédentes. Dambach (1969) rapporte qu'une valeur d'oxygène dissous de 0.8 mg/L fut mesurée en 1929. Dobson et Gilbertson (1971) ont conclu, à partir d'une étude des données antérieures relatives à l'oxygène dissous dans l'hypolimnion du lac Erié, que le taux d'appauvrissement en oxygène de 1970 était plus de deux fois supérieur à l'estimation pour 1929. Ils ont de plus indiqué que la partie centrale du lac Erié était devenue mésotrophe vers 1940 et qu'elle était en train (en 1970) de devenir eutrophe. Cette désoxygénation accrue a été attribuée à une augmentation de la production de phytoplancton découlant d'apports de matières nutritives plus importants

(figure 4). Charlton (1980a) a passé en revue les travaux de Dobson et Gilbertson et a noté l'absence d'une tendance à long terme vers un appauvrissement accru en oxygène. Selon Charlton, les différences entre les taux d'appauvrissement d'une année à l'autre étaient en grande partie dues à des variations de l'épaisseur de l'hypolimnion (figure 5). De fait, les taux actuels d'appauvrissement en oxygène, après correction pour les températures relativement élevées de l'hypolimnion du lac Erié, sont plutôt caractéristiques des lacs mésotrophes (Charlton, 1980a). Charlton (1980b) mentionne dans une publication ultérieure qu'on ne peut comparer la productivité des lacs en termes de concentration d'oxygène à moins de tenir compte de l'épaisseur et de la température de l'hypolimnion (voir les conclusions de Birge et Juday (1911) et de Thienemann (1928) présentées dans l'introduction). Une réévaluation des données de Rosa et Burns (1981), faisant appel à des facteurs de correction pour l'épaisseur de l'hypolimnion, le mélange vertical et les effets thermochimiques dans une zone représentative du bassin central, indique qu'il y a eu, entre 1929 et 1980, un accroissement significatif des taux d'appauvrissement en oxygène.

La Commission mixte internationale (CMI) a passé en revue les rapports annuels de 1973 à 1978 relatifs à la qualité de l'eau des Grands lacs. La situation semble quelque peu confuse. Il est apparu évident, pour la CMI (1975), que l'importance de l'anoxie de l'hypolimnion avait doublé entre 1930 et le milieu des années soixante (tableau 4). Une réduction très importante de l'anoxie a été signalée en 1975, résultat de la formation d'un hypolimnion profond causé par les conditions météorologiques au moment de la formation de la thermocline (CMI, 1976a). Une importante zone anoxique est réapparue en 1976 à la suite de l'appauvrissement, plus rapide qu'en 1975 d'un hypolimnion relativement mince (CMI, 1977). La Commission (CMI, 1977) a conclu que l'importance des vents de printemps et le moment de la première stratification, qui fixent le volume de l'hypolimnion, étaient des facteurs déterminants de

l'étendue des conditions anoxiques. On a noté, en 1977, une réduction de la superficie anoxique, découlant non pas de variations de l'épaisseur de l'hypolimnion, de la concentration de phosphore ou de la production d'algues, mais plutôt d'une modification apportée à la définition du terme anoxie. On considérait, avant 1977, qu'une région était anoxique si la concentration d'oxygène était inférieure à 1.0 mg/L, tandis qu'en 1977, seules les régions où cette teneur était inférieure à 0.5 mg/L étaient considérées comme anoxiques (CMI, 1978a, 1978b).* La CMI a conclu que la région anoxique n'indiquait pas l'état trophique précis du lac Erié (CMI, 1979b).

Un examen de la demande volumique d'oxygène (mg $O_2$/L/d) dans le bassin central, effectué par la CMI (1976a), a montré qu'il y avait eu une augmentation marquée entre 1930 et 1970. La demande d'oxygène avait plus que doublé en quarante ans (tableau 4). Suite aux recommandations pressantes du Plan international de surveillance des Grands lacs (PISGL), une surveillance intensive du lac Erié a été effectuée en 1978 et 1979 afin d'obtenir plus de renseignements sur ces tendances. Les principales conclusions tirées après la première année de ce programme intensif de deux ans étaient que la superficie anoxique du lac Erié ne représentait pas une mesure précise de l'état trophique du lac (CMI, 1979b) à cause de

---

*On définit à l'annexe B (CMI, 1978b) qu'une zone est anoxique quand sa concentration d'oxygène est ≤ 0.5 mg/L. Ce phénomène est illustré à la figure 2.2-8. On définit cependant, dans le haut du tableau 2.2-2 (CMI, 1978b), la zone d'anoxie comme celle où la concentration d'oxygène est ≤ 1.0 mg/L, comme cela est mentionné dans les rapports antérieurs de la CMI. La superficie de la zone d'anoxie de 1977 présentée au tableau, a été calculée à l'aide de la valeur de ≤ 0.5 mg/L, tandis que l'on mentionne, dans le corps du texte que, dans l'hypolimnion de 1977, elle est inférieure à celle signalée en 1976. On ne fait cependant aucune mention du fait que la valeur de 1975, comme toutes les autres valeurs antérieures de superficie, est basée sur le niveau de 1.0 mg/L.

l'influence du niveau des eaux et des conditions météorologiques. De plus, le taux volumique d'épuisement de l'oxygène dans le bassin central n'avait pas varié depuis 1970 (tableau 4), ce qui confirmait la constance du niveau d'eutrophisation générale du lac (CMI, 1979a). La détermination de l'importance des taux volumiques d'appauvrissement en oxygène et de leur relation avec les stratégies de gestion environnementale visant à lutter contre l'eutrophisation sortait du cadre du mandat de la commission (CMI, 1979b). La CMI (1979a) concluait que la faible teneur en oxygène dissous du bassin central du lac Erié résultait d'une augmentation des rejets de phosphore (eutrophisation agricole) ou traduisait une tendance qui n'avait pas changé de façon appréciable au cours des 20 ou 30 dernières années.

On a demandé à un Groupe de travail sur les stratégies de gestion du phosphore d'examiner les divers avis scientifiques et de les commenter. Le Groupe de travail en est venu à la conclusion que le rétablissement de conditions aérobies dans le bassin central était un facteur déterminant et qu'il fallait réduire les concentrations de phosphore pour obtenir des conditions aérobies tout au long de l'année (Thomas et coll., 1980). On a tenté, en faisant appel à l'analyse objective et aux avis d'experts, de déterminer les concentrations maximales de matières nutritives qui permettraient d'obtenir un écosystème de qualité souhaitable dans chacun des Grands lacs. Cinq modèles mathématiques différents ont été élaborés afin d'aider le Groupe de travail à estimer les réponses des lacs à des variations de charge de phosphore (Bierman, 1980; Chapra, 1980; DiToro, 1980; Thomann et Segna, 1980; Vollenweider et coll., 1980). Seulement trois des modèles (Chapra, 1980; DiToro, 1980; Vollenweider et coll., 1980) possédaient des composantes internes traitant de la relation entre les charges de phosphore et l'appauvrissement en oxygène de l'hypolimnion.

Comme l'a souligné Bierman (1980), aucun des modèles étudiés n'avait été testé avec des charges de phosphore autres que les charges actuelles car il n'y avait pas eu de variation appréciable des charges au cours de la période (1967 à 1980) pour laquelle on disposait de données détaillées. Bierman a de plus souligné que les valeurs prévues par les modèles n'étaient que les meilleures estimations possibles et non des prévisions fiables. Pour une charge donnée, les concentrations de phosphore variaient de 5 % à 25 % environ. Il a cependant été intéressant de noter que, même si les réponses aux charges de phosphore des concentrations d'oxygène dissous variaient d'un modèle à l'autre, il n'y avait aucune incohérence fondamentale entre les modèles (Groupe de travail sur les stratégies de gestion, 1980). Etant donné que l'appauvrissement en oxygène est le résultat direct de la concentration de phytoplancton (chlorophylle a) et non de celle du phosphore, il faut aussi souligner que tous les modèles proposés ne permettent qu'une prévision limitée de la teneur en oxygène dissous des eaux hypolimnétiques du lac Erié. Un modèle proposé en 1981 par Vollenweider et Janus a prévu qu'il n'y aurait pas de diminution rapide des taux d'appauvrissement en oxygène du lac Erié tant que la concentration annuelle moyenne de chlorophylle a ne serait pas inférieure à 2 $mg/m^3$. Ce seuil ne saurait être franchi à moins que les rejets de phosphore total dans le lac ne soient inférieurs à 8000 tonnes par an.

Birge et Juday (1911) et Thienemann (1928) ont été les premiers à signaler que l'épaisseur de l'hypolimnion est un paramètre primordial dans le calcul des taux d'appauvrissement en oxygène. Charlton (1979, 1980a et 1980b) s'est récemment intéressé à cette question. Comme nous l'avons déjà indiqué, ce sont les niveaux d'eau et les conditions météorologiques qui déterminent l'épaisseur de l'hypolimnion. La prévision des concentrations d'oxygène dissous est donc, à la limite, fonction des prévisions météorologiques qui, au mieux, ne sont valables qu'à court terme. Les modèles proposés

n'auraient donc que deux possibilités d'utilisation. Un modèle peut être utilisé a posteriori lorsqu'on connait les conditions météorologiques ou il peut être basé sur une condition normale (en donnant une valeur arbitraire à l'épaisseur de l'hypolimnion). Le meilleur commentaire sur les efforts tentés pour modéliser mathématiquement les conditions d'anoxie dans le lac Erié est sans doute celui de Thienemann qui a dit en 1928 que: «Nos tentatives de quantification de phénomènes naturels complexes et en partie biologiques, comme les conditions d'oxygénation des lacs, constituent, dans une certaine mesure, un affront à la nature».

## CONCLUSIONS

Des livres complets ont, au cours de la dernière décennie, été consacrés aux problèmes de l'eutrophisation du lac Erié (Limnological Survey of Eastern and Central Lake Erie, 1928 to 1929; Fish, 1960; Project Hypo, Burns et Ross, 1972a, 1972b; Lake Erie in the Early Seventies, Office des recherches sur les pêcheries du Canada, Vol. 33, numéro spécial). Il est universellement accepté que le lac Erié est eutrophe. Il y a cependant toujours controverse quant à la possibilité de déceler des variations statistiquement significatives de la qualité des eaux. Cette controverse tient au fait qu'on consacre des sommes très importantes aux programmes de surveillance des Grands lacs afin de déterminer les effets de mesures correctives. On a calculé (CMI, 1978a) que les fonds consacrés en 1978 au plan de surveillance des Grands lacs s'élevaient à 9.4 millions de dollars, dont 4.8 millions de dollars ont été consacrés à l'étude intensive du lac Erié (les années 1978 et 1979 sont considérées comme des années d'étude intensive du lac Erié dans le cycle de neuf années du BISGL de la Commission mixte internationale). Le budget du BISGL est établi moitié-moitié entre le Canada et les Etats-Unis. En perspective, le coût du programme de surveillance représente moins de 1 % du montant dépensé annuellement pour l'exécution du programme anti-pollution du bassin des Grands lacs.

Nous avons traité de certains des travaux déjà effectués sur le lac Erié. Notre examen révèle qu'il est extrêmement difficile d'établir pour ce lac des relations de cause à effet ou des tendances statistiquement significatives. Les modifications dans le temps de la communauté biologique peuvent s'expliquer par l'eutrophisation ou l'introduction de produits toxiques (benthos, Groupe de travail sur les stratégies d'élimination du phosphore, 1980), la surexploitation (poisson - Leach et Nepsy, 1976) ou être indicatrices d'effets localisés (phytoplancton - Davies, 1964). L'étude des variations des charges de phosphore total souffre de l'incohérence des méthodes de calcul et de l'insuffisance des données (Zar, 1980), et les concentrations de phosphore dans les eaux du lac sont tellement variables d'une année à l'autre que toute analyse significative semble impossible (Groupe de travail sur les stratégies de gestion du phosphore, 1980). Ou bien les taux d'appauvrissement en oxygène du lac Erié augmentent (Dobson et Gilbertson, 1971) à cause de l'eutrophisation d'origine agricole, ou bien ces taux n'ont pas subi de variations appréciables au cours des 30 ou 40 dernières années (Charlton, 1979, 1980a) et toutes les variations notées traduisent des variations de l'épaisseur de l'hypolimnion. Beeton (1965) a signalé que les modifications de la chimie des eaux du lac Erié résultaient de l'action de l'homme, mais il a trouvé que l'on avait signalé, au cours des 33 dernières années, de faibles concentrations d'oxygène dissous. Pour terminer, on signale que les résultats de la modélisation ne représentent que les meilleures estimations possibles et l'on ne peut garantir le rétablissement des conditions aérobies, ceci même en supposant que les taux des charges respectent les nouveaux objectifs de l'Accord sur la qualité de l'eau des Grands lacs de 1978 (Bierman, 1980; Thomas et coll., 1980).

A remarquer que la CMI n'est pas directement responsable des programmes de recherche dans les Grands lacs. Ces fonctions relèvent de scientifiques des divers organismes gouvernementaux et du domaine universitaire. Le rôle de la CMI se limite à conseiller, à

coordonner et à faire rapport des découvertes. Ainsi, l'étude des divers processus qui affectent l'état trophique du lac Erié doit être sous la responsabilité ultime du monde de la recherche.

Un examen statistique valable des données déjà recueillies pour divers paramètres limnologiques du lac Erié s'impose non seulement pour dégager les tendances qui pourraient exister, mais aussi pour déterminer si les séries de données présentent des faiblesses. Etant donné que le Programme de surveillance des Grands lacs coûte près de dix millions de dollars par année, il est essentiel à ce prix de recueillir le plus d'information possible.

Tableau 1. Données relatives aux Grands lacs[1]

| | Supérieur | Michigan | Huron | Erié | Ontario |
|---|---|---|---|---|---|
| Superficie totale du bassin ($km^2$)[2] | 127 700 | 118 100 | 133 900 | 58 800 | 70 700 |
| Superficie ($km^2$) | 83 300 | 57 850 | 59 510 | 25 820 | 18 760 |
| Volume ($km^3$) | 12 000 | 5 760 | 4 600 | 540 | 1 720 |
| Profondeur moyenne (m) | 145 | 99 | 76 | 21 | 91 |
| Profondeur maximale (m) | 307 | 265 | 223 | 60 | 225 |
| Profondeur moy./ profondeur max. | 0.47 | 0.27 | 0.34 | 0.33 | 0.40 |
| Temps de séjour (années)[3] | 185 | 104 | 27 | 2.7 | 7.8 |

[1] Hutchinson (1957).
[2] Great Lakes Basin Commission. 1976. Limnology of Lakes and Embayments. Great Lakes Basin Framework Study. Annexe n° 4. NOAA.
[3] Rousmaniere. 1979. The Enduring Great Lakes. New York: W.W. Norton and Co.

Tableau 2. Rapport oligotrophie/eutrophie

| Paramètre[1] | Rapport oligotrophie/ eutrophie de Thienemann | Lac Supérieur[2] 27 juil.- 7 août 73 | Lac Erié[2] 28-31 août 73 | Bassin central[2] 28-31 août 73 |
|---|---|---|---|---|
| $\bar{d}$ | 20 | 145 | 21 | 16 |
| % Volume E | 50 | 7 | 48 | 52 |
| $O_2$ H/E | 1.0 | 1.10 | 0.78 | 0.67 |
| Sat. en $O_2$ au fond | 50 | 102.8 | 49.1 | 33.4 |
| $\Delta E$ mL $O_2$/1000 mL | -1.0 | 0.85 | 0.11 | 0.10 |
| $\Delta H$ mL $O_2$/1000 mL | -1.0 | 0.50 | -2.20 | -2.68 |
| $\Delta H$ + E mL $O_2$ | -1.0 | 0.52 | -1.08 | -1.24 |

[1] $\bar{d}$ - profondeur moyenne
%, Volume E - Volume de l'épilimnion exprimé en pourcentage du volume total du lac; l'épilimnion est défini comme la couche s'étendant de 0 à 10 m et l'hypolimnion comme celle s'étendant de 10 m au fond (Thienemann, 1928)
$O_2$ H/E - Coefficient d'oxygène. Concentration d'oxygène de l'hypolimnion divisée par celle de l'épilimnion
Sat. en $O_2$ au fond - Degré de saturation moyenne en oxygène au fond du lac
$\Delta E$ - Différence entre la concentration d'oxygène réelle et la valeur de saturation maximale - épilimnion
$\Delta H$ - Différence entre la concentration d'oxygène réelle et la valeur de saturation maximale - hypolimnion
$\Delta H$ + E - Déficit d'oxygène pour tout le lac
[2] Données tirées d'imprimés des données de surveillance du Centre canadien des eaux intérieures à Burlington (Ontario).
Source : Selon Thienemann (1928).

Tableau 3. Caractéristiques moyennes de l'hypolimnion du bassin central du lac Erié de 1973 à 1980

| Année | Epaisseur (m) | Température (°C) |
|---|---|---|
| 1973 | 4.1 ± 1.0 | 12.0 ± 1.8 |
| 1974 | 5.0 ± 1.0 | 11.5 ± 2.5 |
| 1975 | 7.1 ± 0.6 | 8.1 ± 1.9 |
| 1976 | 4.8 ± 2.6 | 11.6 ± 3.0 |
| 1977 | 4.1 ± 2.1 | 11.1 ± 0.6 |
| 1978 | 5.6 ± 1.2 | 11.6 ± 1.7 |
| 1979 | 4.2 ± 1.5 | 14.1 ± 4.6 |
| 1980 | 5.7 ± 0.5 | 12.8 ± 0.3 |

Données de Fay et Herdendorf (1981).

Tableau 4. Caractéristiques des teneurs en oxygène dissous du bassin central du lac Erié, de 1930 à 1980

| | Superficie estimée de l'hypolimnion anoxique du bassin central du lac Erié | | Demande nette d'oxygène dans le bassin central du lac Erié | | | |
|---|---|---|---|---|---|---|
| Année | Superficie ($km^2$) | % de l'hypolimnion | Taux/unitaire/superficie mg $O_2/m^2$/jour | | Taux/unitaire/volume mg O2/L/jour | |
| 1930 | 300[1] | 3[1] | 0.008[2] | | 0.054[2] | |
| 1940 | | | 0.015[2] | | 0.067[2] | |
| 1950 | | | 0.025[2] | | 0.070[2] | |
| 1959 | 3 600[1] | 33[1] | | | | |
| 1960 | 1 660[1] | 15[1] | 0.037[2] | | 0.093[2] | |
| 1961 | 3 640[1] | 33[1] | | | | |
| 1964 | 5 870[1] | 53[1] | | | | |
| 1967 | 7 500[1] | 68[1] | | | | |
| 1970 | 6 600[1] | 60[1] | 0.039[2] | 0.043[4] | 0.130[2] | |
| 1972 | 7 970[1] | 72.5[1] | | | | |
| 1973 | 11 270[1] | 92.7[1] | 0.023[2] | 0.053[3] | 0.120[2] | 0.120[3] |
| 1974 | 10 250[1] | 87[1] | 0.047[2] | 0.060[3] | 0.110[2] | 0.130[3] |
| 1975 | 400[2] | 4.1[2] | 0.067[2] | 0.067[3] | 0.120[2] | 0.100[3] |
| 1976 | 7 300[3] | 63[3] | | 0.075[3] | | 0.130[3] |
| 1977 | 2 870[4] | 24.8[4] | | 0.058[4] | | 0.130[4] |
| 1978 | 3 980[5] | 71.9[5] | | | | 0.110[5] |
| 1979 | | | | | | |
| 1980 | 4 330[5] | 35.9[5] | | | | |

[1]CNI. 1975. Qualité de l'eau des Grands lacs 1974, Rapport annuel.
[2]CNI. 1975b. Qualité de l'eau des Grands lacs 1975, Annexe B.
[3]CNI. 1977. Qualité de l'eau des Grands lacs 1976, Annexe B.
[4]CNI. 1978a. Qualité de l'eau des Grands lacs 1977, Rapport annuel.
[5]Fay et Herdendorf (1981).

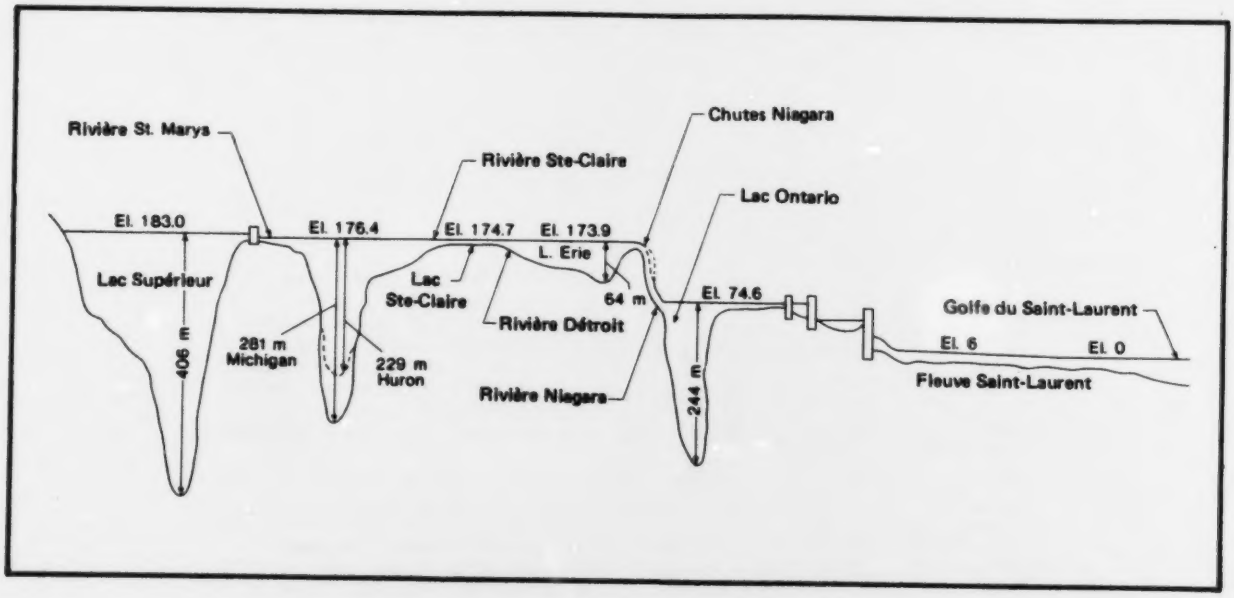



**Figure 1. Les Grands lacs.**

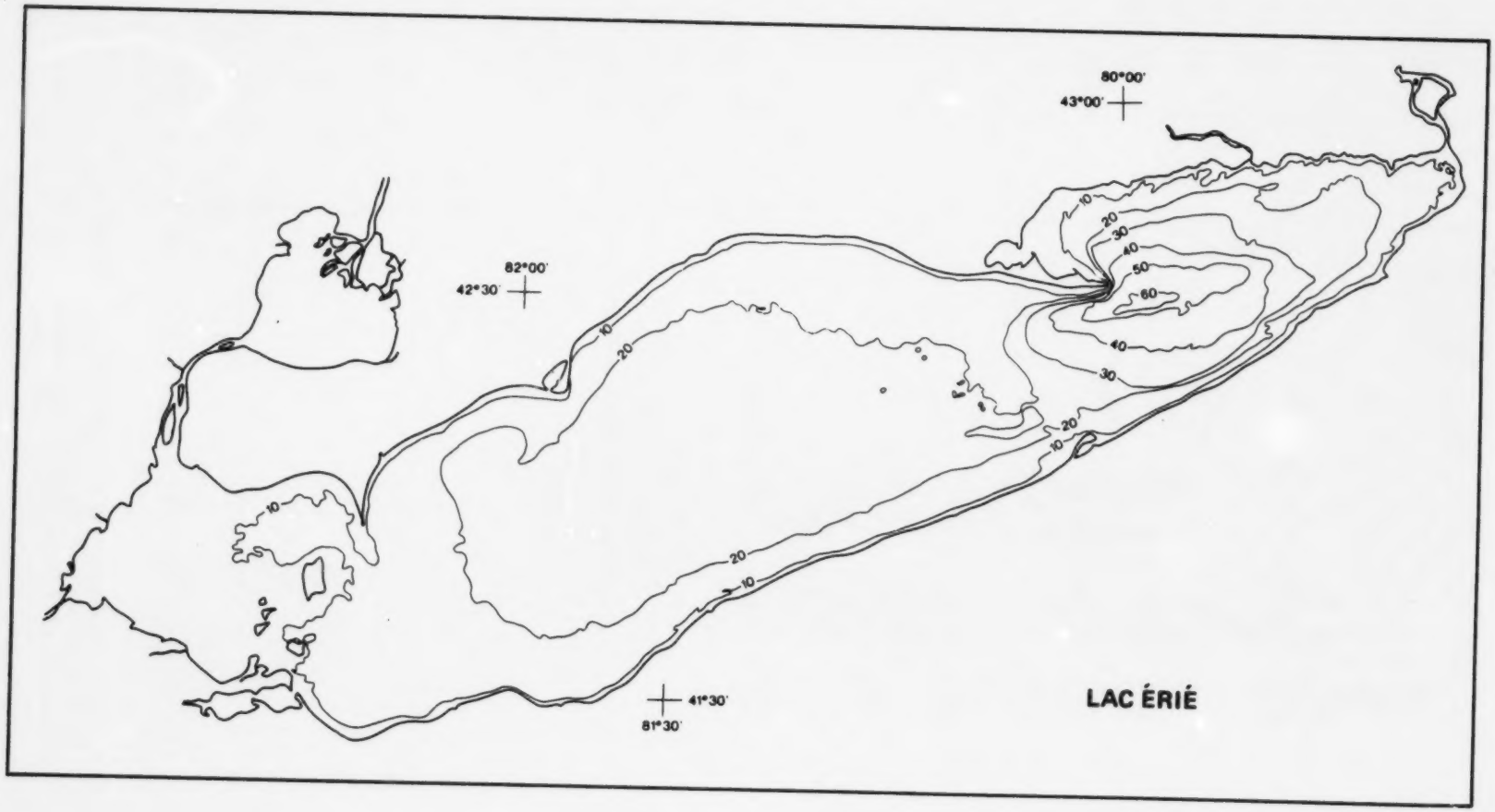



**Figure 2.** Lac Érié, bathymétrie (m).

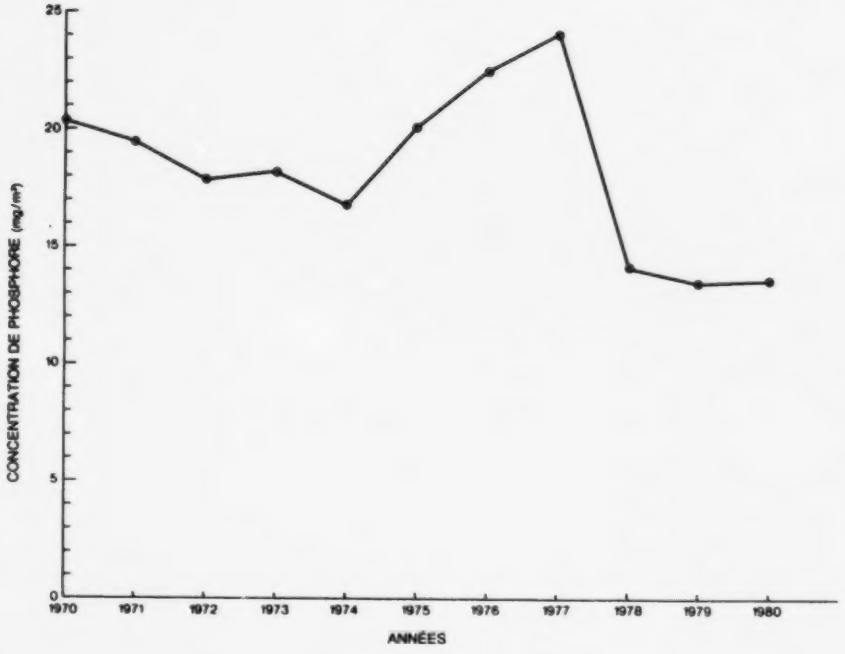



**Figure 3. Concentration de phosphore dans le lac Érié, 1970 à 1980.**

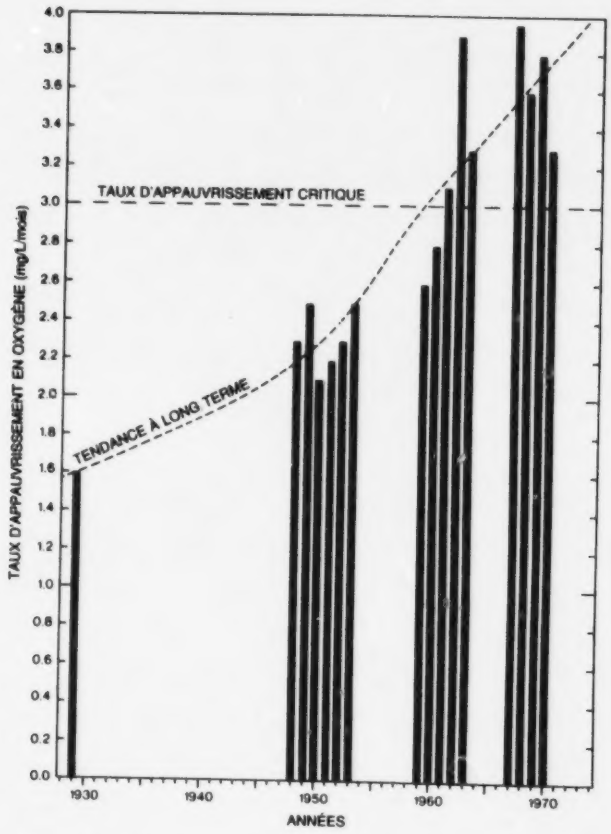



**Figure 4.** Taux moyen d'appauvrissement en oxygène dissous durant l'été dans les eaux de fond, partie centrale du lac Érié. (Gracieuseté de Dobson et Gilbertson, 1971).

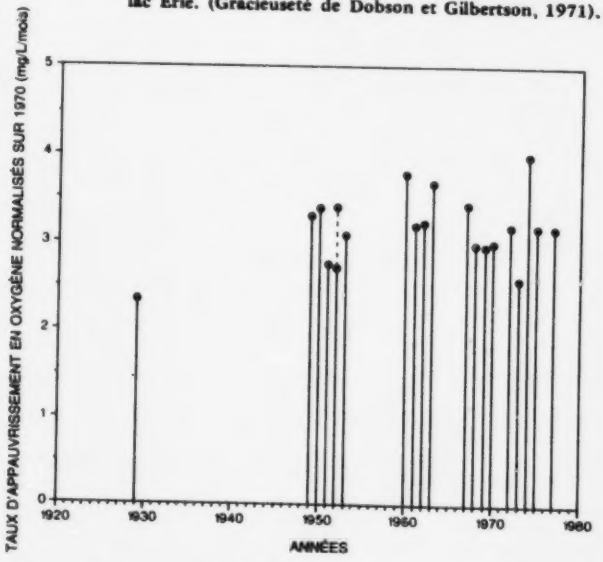



**Figure 5.** Taux chronologique d'appauvrissement en oxygène dissous dans l'hypolimnion du centre du lac Érié, normalisé selon l'épaisseur en 1970. (Gracieuseté de Charlton, 1980a).

# Variations temporelles dans le lac Érié

*par A.H. El-Shaarawi*

## INTRODUCTION

Tout effort pour découvrir l'existence et la structure de tendances que pourrait présenter un indicateur de la qualité de l'eau donné dans un important plan d'eau, comme le lac Erié, se bute à un grand nombre de problèmes. Le premier est la grande variabilité spatiale dans les lacs importants (El-Shaarawi et Shah, 1978; El-Shaarawi et Kwiatkowski, 1977; El-Shaarawi et Esterby, 1981; El-Shaarawi et coll., 1981). Cette variabilité rend difficile la découverte de tendances car elle ajoute une composante supplémentaire à la variance de toute statistique qui pourrait servir à la détection de tendances. De plus, la variabilité spatiale n'est pas constante tout au long de l'année, mais varie avec les saisons (El-Shaarawi, 1982). Le deuxième problème tient au fait que la plupart des variables limnologiques présentent un cycle saisonnier type dont l'allure varie d'une année à l'autre. Si l'on considère une tendance comme l'ensemble des variations d'une année à l'autre, l'analyse des tendances ne se résume donc pas à suivre les variations annuelles en considérant une seule valeur, comme la moyenne, mais à les suivre en comparant une série de courbes représentant chacune le cycle saisonnier d'une année. Le troisième et dernier problème est celui du plan de cueillette des données. Le terme «plan» décrit ici la méthode utilisée pour le choix de l'emplacement des stations d'échantillonnage et du moment de la cueillette des données. L'évaluation adéquate des tendances suppose que les effets spatiaux et saisonniers sont estimés à partir des données et, pour être

adéquates, ces estimations doivent reposer sur un plan d'échantillonnage qui convient à la cueillette des données. Le terme «convenir» est défini en fonction de la quantité d'information obtenue grâce au plan qui convient à la cueillette des données. Ainsi, un plan d'échantillonnage prévoyant des stations fixes pour toute la cueillette des données convient bien à l'évaluation des tendances, tandis qu'un plan prévoyant des stations mobiles convient à la détection des zones où les objectifs de la qualité ne sont pas atteints; enfin, un plan d'échantillonnage combinant des stations mobiles et des stations fixes permet d'obtenir de l'information sur les variations spatiales et temporelles survenant au sein du lac. L'examen du réseau de stations d'échantillonnage utilisé au lac Erié montre que la cueillette des données a été basée sur un plan variable et incohérent. Cela rend difficile la détermination par inférence de variations temporelles de l'état d'eutrophisation du lac. Il est intéressant de souligner ici que l'on a utilisé, depuis 1974, la méthode des stations permanentes pour les échantillonnages effectués au lac Ontario de sorte que l'évaluation des tendances est beaucoup plus facile dans ce cas qu'elle ne l'est pour le lac Erié.

Nous traiterons dans le présent chapitre: (a) des méthodes statistiques qui peuvent être utilisées pour la recherche et la détermination de l'allure de toute tandance que pourraient présenter les valeurs d'un paramètre de la qualité de l'eau et (b) de l'application de ces méthodes à la détermination de tendances des paramètres suivants: (1) niveau de l'eau du lac Erié entre 1900 et 1979, (2) débit de la rivière Niagara entre 1860 et 1975 et (3) température de l'air et de l'eau, profondeur au disque de Secchi, turbidité, teneur en phosphore total, en phosphore réactif soluble, en chlorophylle a et en azote de 1967 à 1980. Une analyse est effectuée pour chacun des bassins (ouest, central et est) du lac. On traitera, dans les deux chapitres suivants, de la tendance du taux d'appauvrissement en oxygène dissous dans l'hypolimnion.

## SOURCES DES DONNEES

Les valeurs annuelles du niveau de l'eau et du débit de la rivière Niagara ont été fournies par la Direction de la planification et de la gestion des eaux du ministère de l'Environnement. Le niveau d'eau est donné par la moyenne des mesures de quatre limnimètres situés à Buffalo, Cleveland, Toledo et Port Stanley. Le débit de la rivière Niagara a été mesuré à Queenston.

Les estimations mensuelles de la température de l'air du bassin central ont été fournies par M. W.M. Schertzer; elles sont aussi présentées dans Lam, Schertzer et Fraser (1983).

Les données relatives aux autres paramètres ont été tirées des dossiers du Centre canadien des eaux intérieures. On a utilisé, pour la présente étude, des teneurs corrigées de chlorophylle a déterminées par analyse d'échantillons d'eau de surface. L'extraction de la chlorophylle a a été effectuée sur filtre de fibre de verre Whatman GF/C; les résultats ont été enregistrés en microgrammes par litre.

Un bathythermogramme électrique en °C a été réalisé pour toute la profondeur du lac avec une sonde Guideline EBT modèle 8031 B et enregistreur HBX-Y. Les températures indiquées par la sonde ont été vérifiées à l'aide d'un thermomètre à renversement.

Les données utilisées pour les paramètres chimiques sont des concentrations dans l'eau de surface. La technique d'analyse des échantillons d'eau est celle décrite dans Analytical Methods Manual (Environnement Canada, 1979) par les chimistes des laboratoires du CCEI, et les résultats sont exprimés en milligrammes par litre.

## METHODES STATISTIQUES

### Réduction de la variabilité spatiale

La stratégie d'échantillonnage utilisée pour obtenir les données du lac Erié n'ayant pas fait appel à des stations fixes, il s'avère impossible de déterminer les variations dans le temps en une ou plusieurs stations données sans éliminer la variabilité spatiale. Il semble approprié de diviser le lac en zones ou régions d'eaux «homogènes» de façon à réduire l'effet de la variabilité spatiale sur la précision des calculs de tendance et l'allure du cycle saisonnier. Cela peut se faire de façon subjective, à partir de connaissances générales de la géomorphologie du lac, ou de façon objective, par une classification comme celle élaborée par El-Shaarawi et Shah (1978). La classification subjective naturelle porte à diviser le lac Erié en trois régions géographiques, soit les bassins est, central et ouest, comme l'indique la figure 6. Les limites du bassin central sont, à l'ouest, une chaine rocheuse et, à l'est, des hauts-fonds de sable et de gravier formant une bande large et peu élevée. L'analyse des tendances sera ici limitée à ces trois zones géographiques. Pour évaluer l'efficacité de tout mode de classification visant à réduire la variabilité spatiale, on peut utiliser la méthode suivante. Soit $x_1$, $x_2 \ldots x_n$ les valeurs observées d'une variable limnologique aléatoire comme la température à n stations d'échantillonnage pendant une mission donnée. La variabilité totale des données est donnée par:

$$Ts = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$= \sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2$$

La méthode utilisée pour la classification aura pour effet de diviser l'ensemble (stations) des données en groupes (zones). Le lac se trouvera donc divisé en k zones. Soit $x_{ij}$ la valeur observée à la $i^{ième}$ station dans la $j^{ième}$ zone ($i = 1,2...n_j$), où $n_j$ est le nombre de stations dans la $j^{ième}$ zone ($j = 1,2...k$) et $\Sigma n_j = n$. De même, la variabilité totale au sein de la $j^{ième}$ zone est:

$$TS_j = \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2, \quad j = 1,2...k$$

où $\bar{x}_j$ est la moyenne des valeurs dans la $j^{ième}$ zone. La variabilité non appliquée par la classification est

$$TS. = TS_1 + TS_2 +...+ TS_k$$

Nous utilisons, comme mesure de l'efficacité de la classification à réduire la variation spatiale totale, la quantité

$$EC = (1 - TS./TS) \times 100$$

La valeur de EC se situe entre 0 et 100. Plus cette valeur se rapproche de 100, plus la classification est efficace. De fait, si on suppose que l'ensemble $x_1$, $x_2...x_n$ a une distribution normale, la variable aléatoire TS./TS a une distribution bêta de paramètres (n-k)/2 et k/2. La moyenne et l'écart-type de EC sont donc, respectivement

$$E[EC] = (k/n)100$$

et

$$\sigma[EC] = (\sqrt{\ } k(n-k)/(n+1)\ /n)100$$

On présente au tableau 5 les valeurs de EC pour les zones géographiques (bassins est, central et ouest), basées sur les températures de surface du lac Erié de 1968 à 1976. Ces valeurs sont aussi reportées à la figure 7 en fonction du mois de l'année. On y note que la valeur de EC croît de mars à juillet et se stabilise à son niveau le plus élevé en août, septembre et octobre avant de décroître en novembre et décembre. La répartition en zones s'avère donc des plus utiles pendant la période allant du milieu de l'été au milieu de l'automne.

Détermination du cycle saisonnier

Une fois le lac divisé en zones, il est possible d'estimer le cycle saisonnier de chaque zone en reliant par régression la variable limnologique (variable dépendante) au temps, en jours (variable indépendante). Les équations de régression ainsi obtenues pour les diverses zones peuvent être soumises à un test d'égalité. Une fois acceptée l'égalité des équations de régression, on peut obtenir un cycle saisonnier unique pour tout le lac. Soit $x_{ijt}$ la valeur de la variable qui nous intéresse dans la $j^{ième}$ région à la $i^{ième}$ station pendant la $t^{ième}$ mission où $j = 1,2...k$; $i = 1,2...n_{jt}$; $t = 1,2...T$; $n_{jt}$ = nombre de stations dans la $j^{ième}$ région pendant la $t^{ième}$ mission; et T = le nombre de missions effectuées pendant l'année de l'étude. Supposons que le cycle saisonnier de la variable limnologique dans la $j^{ième}$ zone puisse être représenté par la fonction $f_j(t)$ comportant p paramètres inconnus, tels que $p \leq k$. Le terme «paramètre» est utilisé ici dans le sens statistique et ne doit pas être confondu avec son usage en limnologie. Ainsi, chlorophylle a est considéré comme un paramètre par les limnologistes, tandis qu'il est considéré ici comme une variable. La pertinence de $f_j(t)$ à représenter le cycle saisonnier peut être testée en définissant

$$PES_j = \sum_{t=1}^{T} \sum_{i=1}^{n_{jt}} (x_{ijt} - \bar{x}_{jt})^2$$

comme la somme des carrés résiduelle «pure» et

$$RSS_j = \sum_{t=1}^{T} \sum_{i=1}^{n_{jt}} (x_{ijt} - \hat{f}_j(t))^2$$

comme la somme des carrés résiduelle du modèle ajusté où $\hat{f}_j(t)$ est l'estimation de $f_j(t)$. L'absence d'ajustement de la somme des carrés est alors définie comme

$$LF_j = RSS_j - PES_j$$

La pertinence de $f_j(t)$ à décrire les variations saisonnières peut être évaluée à l'aide de la statistique

$$F_j = (J_j - T)\ LF_j/(T-p)\ PES_j$$

La statistique $F_j$ a une distribution F à (T-p) et $(J_j - T)$ degrés de liberté où

$$J_j = n_{j1} + n_{j2} + \ldots + n_{jT}$$

Détection des tendances

L'approche utilisée ici pour déceler des tendances consiste à choisir un modèle de cycle saisonnier et à estimer ensuite les paramètres inconnus du modèle pour chaque année. On teste ensuite l'hypothèse selon laquelle ces paramètres ont des mêmes valeurs pour toutes les années. Si le test permet d'accepter l'hypothèse, on conclut à l'absence d'une tendance. Par ailleurs, si l'hypothèse de

l'égalité est rejetée par le test, il faut alors estimer la tendance des variations et en déterminer la relation avec d'autres facteurs.

## NIVEAU DE L'EAU

Le graphique du niveau annuel moyen de l'eau du bassin central du lac Erié est présenté à la figure 8 pour la période allant de 1900 à 1979. Le niveau minimum a été de 568.08 pieds en 1934, et le niveau maximum, de 572.51 pieds en 1974. La fluctuation du niveau annuel moyen a donc été supérieure à quatre pieds au cours de cette période de 80 années. On note dans le graphique que le niveau n'est pas stationnaire, c'est-à-dire que la série des niveaux présente des variations systématiques. Les deux grandes caractéristiques de cette variabilité sont: (i) l'irrégularité des variations cycliques et (ii) la hausse du niveau moyen après 1944. On a, pour lisser les données, calculé la médiane mobile aux cinq ans; les résultats sont présentés à la figure 9 où le niveau de l'eau présente des variations cycliques de fréquence élevée avant 1934 et, après cette date, des cycles de grande amplitude plus espacés.

### Un modèle pour le niveau de l'eau

On a utilisé des méthodes de régression pour élaborer un modèle empirique approprié du niveau de l'eau du lac Erié. L'élaboration du modèle a comporté les étapes suivantes: (i) ajustement d'un premier modèle aux données; (ii) vérification de la pertinence du modèle par examen des résidus et par régression mobile; (iii) modification du modèle en fonction des résultats obtenus en (ii); (iv) ajustement du modèle modifié aux données et répétition des étapes (ii) et (iii); (v) continuation de ce processus jusqu'à l'obtention d'un modèle approprié.

Les méthodes statistiques de base utilisées pour élaborer le modèle reviendront souvent dans le rapport; aussi une description générale de ces techniques est-elle présentée à l'annexe 1.

Le modèle

La représentation graphique indiquant l'existence d'une périodicité des données, il a été décidé de commencer avec un modèle de la forme

$$y_t = \alpha_0 + \alpha_1 \sin \omega t + \alpha_2 \cos \omega t + \varepsilon_t \qquad (3.1)$$

où $y_t$ est le niveau moyen de l'eau de la $t^{ième}$ année ($t = 1,2...80$), $t = 1$ correspondant à l'année 1900 et $t = 80$ à l'année 1979; les paramètres $\alpha_0$, $\alpha_1$, $\alpha_2$, et $\omega$ sont des constantes inconnues; $\varepsilon_t$ est une variable aléatoire de distribution supposée normale, de moyenne 0 et de variance $\sigma^2$ et $\varepsilon_1\ \varepsilon_2 \cdots \varepsilon_{80}$ sont des valeurs indépendantes.

Le modèle 3.1 étant linéaire en $\alpha_0$, $\alpha_1$, et $\alpha_2$, ces derniers paramètres peuvent être facilement estimés par la méthode des moindres carrés. Cependant, le paramètre $\omega$ étant aussi inconnu et non linéaire, l'estimation de ces paramètres ne peut être effectuée que par itération. Supposons tout d'abord que $\omega$ est connu et que $\hat{\alpha}_0(\omega)$, $\hat{\alpha}_1(\omega)$, $\hat{\alpha}_2(\omega)$ et $\hat{\sigma}^2(\omega)$ soient, respectivement, des estimateurs aux moindres carrés de $\alpha_0$, $\alpha_1$, $\alpha_2$ et $\sigma^2$. Il est possible de démontrer que l'estimation de $\omega$ pour $\hat{\omega}$ selon la méthode du maximum de vraisemblance est égale à la valeur de $\omega$ pour laquelle $\hat{\sigma}^2(\omega)$ est un minimum. Il est nécessaire, pour amorcer l'itération, de disposer d'une première valeur pour $\omega$; celle-ci peut être obtenue en divisant le nombre de pics observés dans la série des niveaux d'eau par la longueur de la série et en multipliant le résultat par $2\pi$. On a utilisé la méthode de Newton-Raphson pour déterminer $\hat{\omega}$. Les estimations finales de $\alpha_0$, $\alpha_1$, $\alpha_2$ et $\sigma^2$ sont, respectivement, $\hat{\alpha}_0(\omega)$, $\hat{\alpha}_1(\omega)$, $\hat{\alpha}_2(\omega)$ et $\hat{\sigma}^2(\omega)$. On présente à la figure 10 la courbe de la somme des carrés résiduelle 77 $\sigma^2(\omega)$ en fonction de la période $2\pi/\omega$;

la période estimée du cycle est de 27.5 années. Les niveaux d'eau observés de même que leurs valeurs estimées à l'aide du modèle 3.1 sont reportés en fonction du temps à la figure 11. Ce graphique montre que le modèle initial n'est pas en accord avec les données. A la figure 12, on a reporté les résidus normalisés et orthogonalisés (annexe) en fonction du temps de façon à déceler quelles modifications il y aurait lieu d'apporter au modèle 3.1 pour améliorer l'ajustement. Les tracés des résidus montrent l'existence d'une forte périodicité des données dont ne tient pas compte le modèle. Les résidus orthogonalisés cumulatifs sont aussi reportés en fonction du temps à la figure 13. On y note que le modèle représente bien les niveaux d'eau des 30 premières années; le résidu orthogonalisé cumulatif décroit ensuite de façon constante jusqu'en 1943 avant de croitre à nouveau pendant le reste de la période. Le modèle 3.1 surestime donc les niveaux d'eau pendant la période allant de 1930 à 1943 et il les sous-estime pendant la période de 1944 à 1979. Pour vérifier l'irrégularité de la variance, on a reporté les carrés cumulatifs des résidus, ainsi que leurs valeurs prévues, en fonction du temps; les résultats apparaissent à la figure 14. Il apparait clairement que le modèle n'est pas approprié, car toutes les valeurs observées sont inférieures aux prévisions.

Une technique de régression mobile a aussi été utilisée pour étudier la constance des paramètres de régression en fonction du temps. L'intervalle de temps pour la régression mobile a été fixé à 20 ans, et la valeur de $\omega$, à $\hat{\omega}$. On pourra trouver à la figure 15 la courbe des valeurs estimées de $\hat{\alpha}_0$ en fonction de l'année; les valeurs de $\hat{\alpha}_0$ ne sont pas constantes et peuvent être représentées par une équation quadratique. De façon semblable, les figures 16 et 17 présentent les courbes des valeurs estimées de $\alpha_1$ et $\alpha_2$ en fonction de l'année et on peut y noter l'absence de constance de ces paramètres. Le modèle 3.1 a été modifié pour tenir compte de l'allure de $\alpha_0$

$$y_t = \alpha_0 + \theta_1(t - \bar{t}) + \theta_2 (t - \bar{t})^2 +$$

$$\alpha_1 \sin (2\pi t/27.5) + \alpha_2 \cos (2\pi t/27.5) + \varepsilon_t \qquad (3.2)$$

où $\bar{t} = 40.5$ est la moyenne des nombres 1,2...80 et où $\theta_1$ et $\theta_2$ sont deux autres paramètres inconnus qui reflètent l'allure quadratique de $\alpha_0$. L'effet de l'addition de $\theta_1$ et $\theta_2$ au modèle 3.1 peut être testé à l'aide de la statistique suivante:

$$F = 75 \ (Rés_1 - Rés_2)/2 \ Rés_2$$

où $Rés_1$ et $Rés_2$ sont les sommes des carrés résiduelles obtenues, respectivement, avec les modèles 3.1 et 3.2. Cette statistique a une distribution F à 2 et 75 degrés de liberté. Elle a une valeur calculée de F fortement significative de 17.79 ($p \leq 0.1$).

La répétition de l'analyse précédente en posant le modèle 3.1 comme modèle initial a montré qu'il fallait ajouter une autre composante périodique dont la période était d'environ 40 années. Le modèle 3.2 est donc modifié ainsi:

$$y_t = \alpha_0 + \theta_1(t - \bar{t}) + \theta_2(t - \bar{t})^2 + \alpha_1 \sin(2\pi t/27.5) +$$

$$\alpha_2 \cos(2\pi t/27.5) + \beta_1 \sin(2\pi t/40) + \beta_2 \cos(2\pi t/40) + \varepsilon_t \qquad (3.3)$$

où $\beta_1$ et $\beta_2$ sont introduits pour tenir compte des composantes périodiques additionnelles. L'inclusion de ces deux paramètres améliore l'ajustement de façon appréciable et la statistique correspondante F utilisée pour tester l'importance de $\beta_1$ et $\beta_2$ est:

$$F = 73 \ (Rés_2 - Rés_3)/2 \ Rés_3$$

où $Rés_3$ est la somme des carrés résiduelle du modèle 3.3. La valeur de F est de 5.23 et s'avère significative au niveau de 1 % lorsqu'on la compare à la valeur critique de la distribution F à 2 et 73 degrés de liberté. On trouvera à la figure 18 les niveaux d'eau observés ainsi que les valeurs estimées à partir du modèle 3.3. Un résumé du processus ayant servi à l'élaboration du modèle définitif est présenté au tableau 6.

Variations saisonnières

Les variations saisonnières du niveau de l'eau sont illustrées à la figure 19 où l'on présente les niveaux d'eau moyens au printemps, en été, en automne et en hiver. Les moyennes ont été calculées à l'aide des données de 1965 à 1979. Le niveau de l'eau du lac est le plus élevé en été, puis au printemps, plus faible à l'automne et minimal en hiver. Il apparaît à la figure que les différences entre les valeurs d'été et les valeurs de printemps sont plus importantes entre 1965 et 1973 qu'entre 1974 et 1979.

## DEBIT DE LA RIVIERE NIAGARA

Le débit annuel moyen en milliers de pieds cubes par seconde de la rivière Niagara à Queenston pour la période de 1860 à 1975 est présenté à la figure 20. Il s'agit essentiellement du débit annuel moyen sortant du lac Erié. Cette série chronologique ressemble nettement en gros à celle des niveaux du lac. Cette ressemblance ressort à la figure 21 où les débits de la rivière Niagara sont reportés en fonction des niveaux de l'eau du lac Erié pour la période de 1900 à 1975. On peut y noter que la relation entre ces deux intervalles est linéaire. La droite de régression est aussi présentée à la figure; l'équation de régression est

$$y = 198.132 + 20.677x$$

où y est le débit et x le niveau de l'eau. La valeur de $R^2$ est de 0.97 et l'écart-type de y est estimé à 3.469 pieds cubes par seconde, ceci compte tenu de l'effet du niveau de l'eau. Une élévation du niveau de l'eau du lac d'un pied se traduit donc par une augmentation du débit sortant de 20 677 pieds cubes par seconde.

## TEMPERATURE DE L'AIR

Les températures moyennes mensuelles (avril à novembre) de l'air du bassin central pour la période de 1967 à 1978 sont présentées, en °C, au tableau 7. La structure saisonnière de la température est représentée par la moyenne des moyennes mensuelles (selon le tableau 7) et est donnée à la figure 22a. On peut voir dans cette figure que la température est maximale en août.

Pour étudier les différences d'une année à l'autre dans les températures moyennes mensuelles de l'air (avril à novembre), on a procédé de la façon suivante. Soit $x_{id}$ la température moyenne pour le mois i (i = 1,2...8) pendant l'année d (d = 1,2...12). Soit $\bar{x}_i$ la moyenne des valeurs moyennes du $i^{ième}$ mois

$$\bar{x}_i = \frac{x_{i1} + x_{i2} + \ldots + x_{i12}}{12}$$

L'écart-type entre la moyenne du $i^{ième}$ mois de la $d^{ième}$ année et $\bar{x}_i$ est défini comme

$$R_{id} = x_{id} - \bar{x}_i$$

La structure des résidus $R_{id}$ révèle la nature d'une variation au sein d'une année donnée. Par exemple, on note au tableau 8 pour l'année 1973 des résidus positifs pour tous les mois à l'exception du mois de novembre; cette année peut donc être considérée comme plus

chaude que la moyenne. Les résidus sont négatifs pour l'année 1972, si l'on fait exception du mois de mai; cette année a donc été plus froide que la normale.

Une autre méthode de présentation des différences d'une année à l'autre consiste à construire une courbe des résidus cumulatifs (RC). Le RC est défini par

$$C_{id} = C_{(i-1)d} + R_{id}$$

($i = 1,2...8$) où $C_{id}$ est le RC au $i^{ième}$ mois de la $d^{ième}$ année et $C_{0d} \equiv 0$. Les courbes des RC pour les années 1967 à 1978 sont présentées à la figure 22b et les valeurs de $C_{id}$ sont données au tableau 8. La courbe des RC d'une année normale devrait être très proche de l'axe des x. Elle devrait prendre des valeurs positives les années plus chaudes que la normale et des valeurs négatives dans le cas contraire. Les résultats obtenus pour chacune des années sont résumés ci-après et dans le tableau 9:

1967 - Cette année débute comme une année normale en avril, mai et juin, mais devient plus froide en été et en automne. Comme on peut le voir au tableau 8, la courbe des RC s'approche de l'axe des x en avril, mai et juin, pour ensuite prendre des valeurs décroissantes, ce qui indique que l'été et l'automne ont été plus froids que la moyenne.

1968 - Les résidus du tableau 8 indiquent un printemps normal, un été froid, mais un automne chaud. La courbe des RC révèle que ces variations n'étaient pas très importantes.

1969 - Cette année peut être considérée comme normale. Les résidus ne sont pas très élevés. La température est plus élevée que la normale de juillet à octobre. La courbe des RC ne révèle pas d'écarts importants par rapport à la normale.

1970 - Cette année a été plus chaude que la normale. Ce fait est clairement indiqué par les résidus qui sont, sauf au mois de juillet, tous positifs. La courbe des RC présente une pente croissante, ce qui indique une température annuelle supérieure à la moyenne.

1971 - Les résidus révèlent des valeurs inférieures à la moyenne pour le printemps et l'été, mais supérieures pour l'automne. La courbe des RC n'indique aucun écart important par rapport à la normale.

1972 - Les résidus indiquent une année plus froide que la normale. De fait, seuls les résidus du mois de mai, sont positifs. La courbe des RC le confirme en plongeant sous l'axe des x.

1973 - Les résidus et la courbe des RC indiquent une année plus chaude que la normale.

1974 - A l'exception de ceux d'avril, tous les résidus sont négatifs. Cette année était donc plus froide que la moyenne. La courbe des RC le confirme.

1975 - Les résidus du tableau 8 et la courbe de RC indiquent une température d'été supérieure à la normale.

1976 - Cette année a été plus froide en été et en automne.

1977 - Cette année semble très près de la normale.

1978 - Les résidus indiquent une année plus chaude que la moyenne.

## TEMPERATURE DE L'EAU

### Cycle saisonnier de la température de surface de l'eau

Les cycles saisonniers de la température de surface de l'eau des bassins ouest, central et est du lac Erié sont présentés à la figure 23. Les courbes ont été obtenues à partir des données de températures de surface moyennes recueillies lors des missions de 1967 et 1968. On peut y noter que les températures sont maximales entre les jours 210 et 240, ce qui correspond au mois d'août. La température du bassin ouest est supérieure à celle des bassins central et est jusqu'au jour 260, la tendance étant par la suite inversée.

L'année peut être divisée arbitrairement en trois périodes: (1) le réchauffement, (2) la stagnation et (3) le refroidissement. La période de réchauffement s'étend du jour 1 au jour 200 et celle du refroidissement du jour 250 au jour 365. La période de stagnation d'été correspond au mois de juillet (jour 200 à jour 250), le lac présentant alors une température de surface quasi uniforme. Les températures de surface moyennes du bassin ouest (ordonnée) sont reportées, à la figure 24, en fonction des températures de surface moyennes des bassins central et est; la droite tracée à 45° correspond aux moments d'égale température. Tous les points sont au-dessus de la droite de 45° pendant la période de réchauffement. Cela indique que le bassin ouest est plus chaud que les deux autres; la température la moins élevée a été notée dans le bassin est. Les différences de température entre les trois bassins diminuent cependant avec le temps, donc avec la température. Pendant la période de refroidissement, le bassin le plus froid est le bassin ouest et les différences entre les trois bassins augmentent à mesure que la température décroit, donc avec le temps. On peut aussi noter sur ce graphique que la différence maximale des températures moyennes entre le bassin ouest et le bassin central, pendant la période de

réchauffement, a été de 8° environ; la température du bassin ouest était alors de 12°C environ. Pendant la période de refroidissement, la différence maximale de température a été d'environ 4°C, la température de surface du bassin ouest étant aussi d'environ 4°C.

Analyse statistique des données de température de surface

L'analyse statistique qui suit vise l'élaboration de méthodes statistiques permettant d'estimer le cycle annuel des températures saisonnières et d'en déterminer les variations d'une année à l'autre. Soit $y_{ts}$ la moyenne des températures de surface observées le $s^{ième}$ jour du calendrier ($s = 1,2...365$) de la $t^{ième}$ année ($t = 1,2...m$) et m le nombre d'années de la période étudiée. On suppose que le cycle saisonnier de la température pour la $t^{ième}$ année peut être représenté par le modèle

$$y_{ts} = \alpha_t - \beta_t \cos ws + \gamma_t \sin ws - \epsilon_{ts} \qquad (3.4)$$

où $w = 2\pi/365$; $\alpha_t$, $\beta_t$ $\gamma_t$ sont des paramètres inconnus et $\epsilon_{ts}$ est une variable aléatoire normale de moyenne 0 et de variance $\alpha_t^2/n_{ts}$, où $\sigma_t^2$ est la variance d'une mesure unique de température et $n_{ts}$ le nombre de mesures de la température utilisées pour calculer la moyenne $y_{ts}$.

Ce modèle a été retenu parce que le cycle saisonnier observé des températures (figure 25) a l'allure d'un cosinus. Le modèle 3.4 peut aussi être représenté de la façon suivante:

$$y_{ts} = \alpha_t + R_t \cos\left(\frac{2\pi s}{365} + \phi_t\right) + \epsilon_{ts} \qquad (3.5)$$

où l'amplitude $R_t$ et la phase $\phi_t$ sont obtenues à partir de termes $\beta_t$ et $\gamma_t$ par les relations:

$$R_t = \sqrt{\beta_t^2 + \gamma_t^2}$$

et

$$\tan \phi_t = \gamma_t / \beta_t$$

L'expression $\alpha_t + R_t$ représente la température maximale estimée pour la $t^{ième}$ année, tandis que l'expression $D_t = -365\ \phi_t / 2\pi$ représente la date où la température maximale a été notée. Le modèle 3.4 s'avère préférable au modèle 3.5 pour l'analyse statistique, car il est linéaire en $\alpha_t$, $\beta_t$, et $\gamma_t$, paramètres inconnus qui peuvent donc être estimés par application directe de la méthode des moindres carrés. Les paramètres $R_t$ et $\phi_t$ du modèle 3.5 sont cependant plus significatifs. Le tableau 10 résume les résultats de l'ajustement du modèle 3.4, et donc du modèle 3.5, aux températures de surface moyennes pour chacun des bassins du lac Erié pendant les années 1967 à 1978. Les résultats pour chacun des bassins comprennent l'estimation de la température maximale de chaque année, la phase $\phi_t$ estimée, le nombre estimé de jours du calendrier $D_t$ nécessaire pour atteindre la température maximale, le coefficient de détermination $R_t^2$ et la statistique F pour le test de signification de la régression. Toutes les valeurs de F sont marquées par deux astérisques (**) afin d'indiquer que les équations de régression sont significatives au niveau de 1 %. De plus, la plus petite valeur obtenue pour le coefficient de détermination $R^2$ est 0.76. La température de surface maximale estimée de l'eau du bassin ouest a varié de 21.47°C, en 1971 à 25.60°C, en 1973. Les valeurs correspondantes du bassin central vont de 19.46°C, en 1971, à 23.90°C, en 1973. Ces valeurs sont en accord avec les résultats déjà signalés pour la température moyenne de l'air en 1973 et 1975 au bassin central. Dans le bassin est, la température maximale estimée est la moins élevée (19.78°C) en 1969 et la plus élevée (23.78°C) encore en 1973.

Les valeurs de $D_t$ du tableau 10 indiquent aussi une importante variation. Dans le bassin ouest, le plus petit nombre de jours $D_t$ nécessaire pour atteindre la température maximale a été de 203.2 jours en 1971, et le plus grand de 221.4 jours en 1969. Les valeurs correspondantes de $D_t$ minimum et maximum pour le bassin central ont été respectivement de 206.8 en 1967 et de 229 en 1969. Par ailleurs, les valeurs correspondantes de $D_t$ pour le bassin est ont été de 209.2 en 1971 et de 232.8 en 1968. Les valeurs médianes de $D_t$ ont été de 217.2, 226.1 et 226.6 jours respectivement dans les bassins ouest, central et est. On peut donc conclure que la valeur maximale est atteinte dans le bassin ouest 9 jours avant qu'elle ne le soit dans le bassin central. La différence entre les médianes $D_t$ des bassins central et est est négligeable.

Variations d'une année à l'autre de la température de surface de l'eau

Le modèle 3.4 peut servir à déterminer si la variabilité d'une année à l'autre du cycle saisonnier des températures de surface est supérieure à celle prévue statistiquement. Cela équivaut à tester l'hypothèse

$$H: \alpha_t = \alpha,\ \beta_t = \beta \text{ et } \gamma_t = \gamma$$

pour toutes les valeurs de t. Ainsi, on suppose que les paramètres du modèle 3.4 sont indépendants de l'année et qu'il est possible d'établir une seule équation pour représenter adéquatement la température, ce qui est facilement réalisable en utilisant la statistique F de Fisher pour tester l'égalité de plusieurs équations de régression. Etant donné l'inégalité des valeurs de $\hat{\sigma}_t^2$, l'analyse a été effectuée avec des poids empiriques $\hat{\sigma}_t^2$ qui sont les valeurs estimées de $\sigma_t^2$. Les résultats de cette analyse sont résumés au tableau 11 où on peut noter que toutes les valeurs de F de chaque bassin sont significatives au niveau de 1 %.

Pour évaluer la structure et l'importance générale de la variabilité d'une année à l'autre, on a reporté aux figures 26a, 26b, et 26c les résidus moyens annuels en fonction des années pour les bassins ouest, central et est respectivement. Les résidus sont définis comme les différences entre les valeurs observées et les valeurs estimées à l'aide du modèle 3.4 dont on a ensuite fait la moyenne pour chaque année. Un résidu moyen positif indique une année plus chaude que la normale, une année plus froide étant indiquée par une valeur négative. On trouve à l'examen des graphiques que les années 1972 et 1975 ont été les plus froides et que les années 1971 et 1973 ont été les plus chaudes.

Variation verticale de la température

Les eaux du lac Erié peuvent, pendant l'été, être divisées en trois couches distinctes selon leur température et, conséquemment, leur densité. Ces couches sont: l'épilimnion, la thermocline ou mésolimnion et l'hypolimnion. Les profils bathygraphiques de température obtenus lors de la mission effectuée du 19 juillet au 3 août 1968 sont présentés à la figure 27. L'épilimnion est la couche supérieure la moins dense dans laquelle la température de l'eau varie de façon graduelle. On trouve, en dessous, la thermocline dont les gradients de température et de densité sont marqués. La couche la plus profonde, l'hypolimnion, est celle où l'eau est la plus dense, et la température, constante. L'hypolimnion du bassin central se trouve à une profondeur moyenne d'environ 4 m mais cette valeur varie tant dans l'espace que dans le temps. Elle est beaucoup plus profonde dans le bassin est à cause de la plus grande profondeur de ce dernier et elle est inexistante dans le bassin ouest. Cette configuration est inversée durant l'hiver. La masse d'eau la plus froide se trouvant au fond du lac l'été et en surface l'hiver, il doit y avoir un moment, entre le milieu et la fin de l'automne, où tout le lac devient isotherme. Ce phénomène est désigné sous le nom d'inversion d'automne. De façon similaire, la période de stagnation d'hiver est suivie de l'inversion de printemps.

On trouvera aux figures 28 et 29 les profils bathygraphiques de température pour chaque station du bassin central ayant fait l'objet d'un relevé en 1967 et 1978 respectivement. On peut faire deux observations: (i) l'épilimnion, l'hypolimnion et la thermocline ne sont pas nettement définis et (ii) les limites de ces zones ne sont pas les mêmes en 1967 et en 1978. Ce décalage des limites des couches s'explique par des différences entre les structures de la température au cours de ces deux années. Comme cela a été montré plus haut, l'année 1967 a été plus froide et l'année 1978 plus chaude que la moyenne. De telles différences sont importantes lorsque l'on traite de modifications de la qualité des eaux.

Variabilité d'une année à l'autre de la température de l'hypolimnion (bassin central)

On présente à la figure 30 les températures moyennes observées dans l'épilimnion et l'hypolimnion du bassin central. La période de stratification du bassin central varie de 150 à 250 jours. L'épilimnion et l'hypolimnion sont nettement séparés. La température moyenne minimale de l'hypolimnion était d'environ 6°C, et la température moyenne maximale légèrement supérieure à 13°C. On a aussi noté que les températures de l'hypolimnion augmentaient légèrement en fonction du temps. Les valeurs de température ont été ajustées au modèle de régression linéaire

$$y_t = \alpha + \beta t + \varepsilon_t$$

où $y_t$ est la température moyenne du bassin central au $t^{\text{ième}}$ jour du calendrier, $\alpha$ et $\beta$ sont, respectivement, la pente et l'ordonnée à l'origine de la droite, et $\varepsilon_t$ est une variable aléatoire normale de moyenne 0 et de variance $\sigma^2$. On trouvera, pour chaque année, les équations de régression ajustées à la figure 30 et les valeurs estimées de $\alpha$ et $\beta$ au tableau 12. On peut voir dans la figure que les températures d'hypolimnion les plus faibles

ont été notées, dans l'ordre, en 1975 et en 1978. La température la plus élevée a été notée en 1977. De plus, la pente estimée, $\beta$, renseigne sur le transfert de chaleur de l'hypolimnion vers les autres couches. Les valeurs de $\beta$ de 1967, 1968 et 1973 sont très proche de 0, ce qui indique un transfert de chaleur minimal pendant ces années. Le transfert de chaleur a été assez faible ($0.025 < \beta < 0.05$) en 1969, 1971, 1978 et 1979, tandis qu'il a été plus élevé ($0.05 < \beta < 0.09$) lors des autres années.

## CHLOROPHYLLE a

On présente à la figure 31 les courbes des concentrations moyennes non corrigées de chlorophylle a en surface en fonction des jours du calendrier pour les bassins ouest, central et est du lac Erié. Les données pour chaque bassin ont été recueillies au cours des années suivantes: 1968, 1970 et 1972 pour le bassin ouest; 1968, 1970, 1979 et 1980 pour le bassin central et 1968, 1970, 1972 et 1980 pour le bassin est. On note dans la figure que la biomasse du phytoplancton était la plus abondante dans le bassin ouest et la plus rare dans le bassin est. L'écart de biomasse était beaucoup plus important entre le bassin ouest et le bassin central qu'entre ce dernier et le bassin est. La chlorophylle a présente les pics les plus élevés en 1970, tandis que le cycle saisonnier de 1980 est moins marqué que ceux des autres années.

### Bassin ouest

Les résultats obtenus suite à l'ajustement du modèle 3.4 aux données du bassin ouest sont présentés au tableau 13 où l'on peut noter que le coefficient de détermination est relativement élevé (0.74). Une tendance à la hausse des valeurs du maximum estimé apparait de façon évidente. Pour vérifier si la chlorophylle a présente le même cycle saisonnier avec des conditions initiales différentes, on a raffiné le modèle 3.4:

$$y_{ts} = \alpha_t + \beta \cos wt + \sin wt + \epsilon_{ts} \qquad (3.6)$$

l'hypothèse testée est donc

$$H_1: \beta_t = \beta \text{ et } \gamma_t = \gamma \text{ pour toutes les valeurs de t}$$

Si $H_1$ est acceptée, cela signifie que le modèle 3.6 est suffisamment ajusté aux données et la prochaine hypothèse à tester est

$$H_2: \alpha_t = \alpha \text{ pour toutes les valeurs de t}$$

Si $H_2$ est aussi acceptée, on peut alors conclure qu'il n'y a pas de différence d'une année à l'autre. Les statistiques $F_1$ et $F_2$ sont utilisées pour tester les hypothèses $H_1$ et $H_2$:

$$F_1 = (n - 3m)\ (RES_1 - RES)/2(m-1)RES$$

et (3.7)

$$F_2 = (n-m-2)\ (RES_2 - RES_1)/(m-1)RES_1$$

Les quantités n et m représentent, respectivement, le nombre total de missions et le nombre d'années; RES, $RES_1$ et $RES_2$ représentent respectivement, la somme des carrés résiduelle selon le modèle 3.6, l'hypothèse $H_1$ et l'hypothèse $H_2$. Ces modèles supposent que $F_1$ et $F_2$ ont des distributions F à 2(m-1), (n-3m) et (m-1) et (n-m-2) degrés de liberté respectivement. Il est aussi possible de tester l'hypothèse voulant que toutes les années présentent le même cycle saisonnier. L'hypothèse testée est

$$H_0: \alpha_t = \alpha,\ \beta_t = \beta \text{ et } \gamma_t = \gamma \text{ pour toutes les valeurs de t}$$

La statistique $F_0$ utilisée pour le test $H_0$ est donnée par l'équation

$$F_0 = (n-3m)\ (RES_0 - RES)/3(m-1)RES,$$

où $RES_0$ est la somme des carrés résiduelle selon $H_0$.

Les valeurs des statistiques $F_1$, $F_2$ et $F_0$, et leurs degrés de liberté, sont donnés au tableau 14. Aucun de ces tests ne révèle de différences statistiquement significatives d'une année à l'autre, mais les valeurs de $\hat{\alpha}_t$ obtenues à l'aide du modèle 3.6 indiquent, comme on peut le voir au tableau 15, une augmentation systématique de la teneur de chlorophylle a dans le bassin ouest.

Bassin central

L'analyse précédente a été répétée pour le bassin central. Les résultats obtenus sont présentés aux tableaux 16, 17 et 18. On voit au tableau 17 que l'hypothèse $H_1$ n'était pas significative. Le fait que $H_2$ était significative au niveau de 5 % indique cependant qu'il y a des différences d'une année à l'autre, donc que le modèle 3.6 peut être utilisé pour représenter la chlorophylle a dans le bassin central. Les valeurs de $\hat{\alpha}_t$ indiquent que la chlorophylle a était la plus abondante en 1970 et la moins abondante en 1980. Il semble aussi qu'il y ait eu, depuis 1970, une tendance linéaire à la baisse de la chlorophylle a.

Bassin est

On trouvera aux tableaux 19, 20 et 21 les résultats de l'analyse statistique portant sur la bassin est. Le modèle 3.6 semble convenir à la description de la chlorophylle a dans le bassin est, du fait que la valeur de $F_1$ n'est pas significative au niveau de 5 %, mais que celle de $F_2$ l'est au niveau de 1 %, cela montre que $\beta_t$ et $\gamma_t$ demeurent les mêmes d'une année à l'autre et que la variation de $\alpha_t$ est statistiquement significative. Les valeurs de $\hat{\alpha}_t$ sont présentées au tableau 21 et indiquent une tendance à la baisse de la chlorophylle a.

En résumé, il semble y avoir une tendance à la baisse significative de la chlorophylle a dans les bassins central et est entre 1968 et 1980 et une tendance à la hausse non significative dans le bassin ouest entre 1968 et 1972. Ces résultats sont présentés à la figure 32 où l'effet des années, $\hat{\alpha}_t$, est reporté en fonction des années pour chacun des trois bassins.

## PHOSPHORE TOTAL (PT)

Les teneurs moyennes en phosphore total (PT) sont reportées à la figure 33 en fonction du jour de l'année, ceci pour les eaux de surface des trois bassins. L'allure générale du cycle saisonnier du PT révèle que les valeurs sont élevées en hiver et au printemps pour ensuite diminuer rapidement au début de l'été et atteindre leur minimum à la fin de l'été. Les teneurs en PT sont aussi plus élevées dans le bassin ouest, comparativement aux deux autres bassins. L'importance de la variabilité d'une année à l'autre semble ne pas être la même pour tous les bassins. La figure montre que la variabilité est la plus élevée dans le bassin ouest, puis dans le bassin central. De plus, le cycle saisonnier peut être représenté par le modèle 3.4 ou par une fonction quadratique. On a décidé de reprendre l'analyse qui avait été faite pour la chlorophylle a.

### Bassin ouest

Les résultats de l'analyse statistique des teneurs en PT effectuée à l'aide du modèle 3.4 sont présentés aux tableaux 22, 23 et 24. On peut voir au tableau 23 que le modèle 3.6 donne une représentation raisonnable des données. La valeur de $F_2$ est très significative, ce qui indique que la variabilité d'une année à l'autre, liée à celle de $\alpha_t$, est appréciable. On note aussi au tableau 24 que les valeurs du PT de 1977 et 1978 sont beaucoup moins importantes que celles de 1968 à 1971. On peut même conclure à une légère diminution de la valeur du PT quand celle de 1968 est comparée

à celles de 1970 et de 1971. De plus, il y a eu un net déclin de cette valeur en 1972.

Bassin central

Les résultats, présentés aux tableaux 25, 26 et 27, indiquent que le modèle 3.6 est approprié. Les valeurs de $\hat{\alpha}_t$ indiquent l'existence d'une tendance à la baisse significative du PT.

Bassin est

Le PT du bassin est tend à la baisse de façon significative. Les variations sont cependant beaucoup moins prononcées que celles notées dans les bassins ouest et central. Les résultats présentés aux tableaux 28, 29 et 30 montrent l'absence de variations statistiquement significatives au niveau de 1 %. Les valeurs de $\hat{\alpha}_t$ du tableau 30 indiquent une diminution après 1971.

On a reporté à la figure 34 les valeurs de $\hat{\alpha}_t$ en fonction de l'année, ceci pour chacun des trois bassins. Les valeurs du PT y diminuent de façon appréciable après 1971; l'importance de cette diminution décroit dans l'ordre des bassins ouest, central et est. De fait, la concentration du PT du bassin ouest se rapproche de beaucoup de celles des bassins central et est en 1977, 1978, 1979 et 1980. On présente à la figure 35 les valeurs de $\hat{\alpha}_t$ relatives à la concentration de chlorophylle a en fonction des valeurs correspondantes du PT. Le graphique montre une forte association positive entre ces deux paramètres: la diminution du PT est associée à celle de la chlorophylle a.

## PHOSPHORE REACTIF SOLUBLE (PRS)

Le traitement des résultats étant très semblable au précédent, on a décidé de les présenter simultanément pour les trois

bassins, et non d'en faire une description complète pour chacun des cas. Les valeurs de $F_2$ et de $F_0$ (tableau 31) sont, pour les bassins ouest et central, significatives au niveau de 1 % et, pour le bassin est, significatives au niveau de 5 %. La valeur de $F_1$ n'étant pas significative pour les bassins ouest et est et n'étant significative qu'au niveau de 5 % pour le bassin central, on peut conclure que le modèle 3.6 convient à la description du PRS. Les valeurs estimées de $\alpha_t$ sont présentées au tableau 32. On y remarque que la concentration du PRS a diminué de façon appréciable après 1977, comparativement à sa valeur d'avant 1973. On pourra trouver à la figure 36 les valeurs de $\hat{\alpha}_t$ reportées en fonction des années pour chacun des trois bassins.

## NITRATE ET NITRITE FILTRES ($N\text{-}NO_3NO_2$)

On présente à la figure 37 les tracés des concentrations de N sous forme de $NO_3$ et de $NO_2$ dans le bassin ouest en fonction du jour de l'année, ceci pour chacune des années pour lesquelles des données existent. Les teneurs en $N\text{-}NO_3NO_2$ atteignent un minimum entre les jours 220 et 280, au cours du mois d'août. Le modèle 3.4 est ajusté aux données et les valeurs estimées de la teneur en $N\text{-}NO_3NO_2$ ainsi que l'amplitude sont représentées par $\hat{\alpha}_t$ et $R_t$ respectivement au tableau 33 où $R_t = \sqrt{\beta_t^2 + \gamma_t^2}$. Ces valeurs sont données pour chacun des bassins et pour chacune des années pour lesquelles il existe des données. Les valeurs de $\hat{\alpha}_t$ et de $R_t$ obtenues pour le bassin ouest sont supérieures à celles des bassins central et est, ce qui indique que la valeur et l'amplitude des concentrations de $N\text{-}NO_3NO_2$ sont les plus élevées dans le bassin ouest. De plus, les valeurs de $\hat{\alpha}_t$ sont minimales en 1971 et maximales en 1978 et 1980. C'est aussi dans le bassin ouest que le taux d'augmentation de la teneur en $N\text{-}NO_3\text{-}NO_2$ est le plus faible. Les valeurs de $F_2$ et de $F_0$ de chacun des trois bassins sont présentées au tableau 34. Elles sont toutes significatives au niveau de 1 %, l'hypothèse $H_0$ est donc

rejetée. $F_2$ étant significatif, le modèle 3.6 ne s'avère donc pas approprié et c'est le modèle 3.4 qui représente de façon adéquate les variations d'amplitude du cycle saisonnier. Les valeurs de $N\text{-}NO_3NO_2$ observées et estimées à l'aide du modèle 4 pour le bassin ouest sont présentées à la figure 37. On peut y noter que les cycles saisonniers du $N\text{-}NO_3NO_2$ sont les mêmes de 1967 à 1970 et que le cycle de 1978 est similaire à celui de 1980. Cette hypothèse est ensuite testée par la statistique F appropriée et est acceptée au niveau de 5 %. Un résumé des paramètres du modèle modifié est présenté au tableau 35. On peut y voir que les valeurs de $N\text{-}NO_3NO_2$ ont été plus élevées en 1978 et 1980 que pendant les autres années et que la valeur la plus faible a été notée en 1971. On pourra trouver aux figures 38a, 38b et 38c les teneurs en $N\text{-}NO_3NO_2$ observées et estimées pour le bassin ouest: les résultats obtenus pour les années 1967, 1968 et 1970 apparaissent à la figure 38a, ceux de 1971 à la figure 38b et ceux de 1978 à la figure 38c. Des données semblables sont présentées aux figures 39a, 39b et 39c pour le bassin central et on trouvera à la figure 40 les teneurs en $N\text{-}NO_3NO_2$ observées et estimées (modèle 3.4) pour le bassin est.

## AMMONIAC ($NH_3$)

L'analyse précédente a été appliquée aux valeurs de $NH_3$. Les résultats obtenus ont montré que le modèle 3.6 convient aux bassins ouest et est, mais les données relatives au bassin central ont indiqué une petite variation de l'amplitude, de sorte que le modèle 3.4 s'avérait plus approprié. Les valeurs de $F_1$, $F_2$ et $F_0$ utilisées pour tester les hypothèses $H_1$, $H_2$ et $H_0$ respectivement sont présentées au tableau 36. Les résultats obtenus montrent que l'hypothèse $H_2$ est fortement significative si l'on fait exception du bassin ouest, ce qui indique qu'il y a des différences d'une année à l'autre. Les valeurs de $NH_3$ observées et estimées pour les bassins ouest, central et est sont respectivement

présentées aux figures 41, 42 et 43. On peut y voir que le modèle 3.6 représente bien les données et qu'il y a tendance à la baisse des teneurs en $NH_3$.

Les valeurs estimées du paramètres du modèle 3.6 sont présentées au tableau 37. Les valeurs de $\hat{\alpha}_t$ sont plus importantes pour le bassin central que pour les deux autres bassins. On y note aussi une diminution appréciable après 1977, comparativement à 1968, 1970 et 1971.

## PROFONDEUR AU DISQUE DE SECCHI

Un examen préliminaire de la relation entre la moyenne et l'écart-type des profondeurs mesurées au disque de Secchi a montré la nécessité d'une transformation logarithmique des données avant leur analyse. Les résultats de l'ajustement du modèle 3.4 aux données logarithmiques sont présentées au tableau 38. Les profondeurs augmentent avec le temps pour atteindre un maximum en juillet.

Les valeurs présentées sous l'en-tête amplitude au tableau 38 sont les logarithmes des profondeurs maximales estimées au disque de Secchi et celles sous $D_t$ sont les estimations du nombre de jours nécessaires pour atteindre ces maximums. Comme prévu, l'amplitude la plus faible est celle du bassin ouest, suivie de celles des bassins central et est. Le bassin ouest est aussi celui qui présente le cycle saisonnier le moins bien défini; cela se dénote par les valeurs du coefficient de détermination $R^2$ ou les valeurs de la statistique F qui mesurent la signification de la régression. Les valeurs $R^2$ les plus faibles sont celles du bassin ouest et la plupart des valeurs de F ne sont pas significatives. Cela contraste avec les valeurs correspondantes des deux autres bassins. On présente à la figure 44 les courbes des logarithmes des profondeurs maximales au disque de Secchi pour chacun des bassins. On peut y noter que la courbe du bassin ouest est bien distincte des deux autres ainsi qu'une augmentation des valeurs en 1976, 1977 et 1978.

Les estimations du nombre de jours $D_t$ nécessaire pour atteindre la valeur maximale présentent le même genre de variabilité. A l'exception de 1977, cette valeur maximale varie entre 188 et 240 jours, et elle est atteinte dans le bassin central avant de l'être dans le bassin est.

Variabilité d'une année à l'autre des profondeurs au disque de Secchi

La statistique F utilisée pour tester la constance de l'équation de régression d'une année à l'autre est présentée au tableau 39. Elle est significative au niveau de 1 % pour les bassins central et est, mais non pour le bassin ouest. Comme cela a été précédemment noté, il semble y avoir une augmentation des profondeurs maximales au disque de Secchi, estimations qui renseignent sur le type de variation d'une année à l'autre.

Un autre type de variabilité d'une année à l'autre peut être déterminé à partir du graphique des résidus moyens obtenus après ajustement d'une équation unique à toutes les données. Les courbes des données et des valeurs estimées à partir du modèle 3.4 sont présentées en fonction du temps à la figure 45. Celles des résidus moyens sont présentées à la figure 46. Il apparaît clairement que la figure 46 ne présente aucune variation importante si ce n'est, peut-être, une baisse des valeurs de 1973.

## TURBIDITE

Les valeurs de turbidité ont aussi subi une transformation logarithmique avant l'ajustement du modèle 3.4. Les résultats sont présentés pour les années 1967 à 1972, années pour lesquelles des données de turbidité étaient disponibles.

Analyse des résultats de l'ajustement du modèle 3.4 aux données logarithmiques de turbidité

Ces résultats sont résumés au tableau 40. La structure du cycle saisonnier est très bien expliquée par le modèle 3.4. Les valeurs de la statistique F sont fortement significatives et les valeurs correspondantes de $R^2$ sont élevées. La seule exception tient aux données de 1972 pour les bassins central et est. Les valeurs de turbidité sont minimales vers la fin de juillet et le début d'août et le minimum est d'abord atteint dans le bassin ouest. La structure générale indique une augmentation de la turbidité dans tous les bassins entre 1969 et 1972. Celle-ci est beaucoup plus prononcée si on la compare aux résultats déjà obtenus pour la profondeur au disque de Secchi.

Variabilité de la turbidité d'une année à l'autre

Les valeurs de la statistique F relatives aux différences d'une année à l'autre sont présentées au tableau 41. Toutes ces valeurs sont fortement significatives (au niveau de 1 %) et dénotent donc l'existence de différences entre les cycles saisonniers d'une année à l'autre.

## RELATION PHOSPHORE - CHLOROPHYLLE a

Il est bien connu que l'eutrophisation des lacs a pour cause des apports excessifs de phosphore et d'azote (Vollenweider et coll., 1980). L'apport de phosphore dans le lac Erié par des sources ponctuelles est passé d'environ 10 000 tonnes métriques par an, en 1972-1973, à 5700 tonnes métriques par an en 1977 (Slater et Bangay, 1980). Il s'avère donc important de déterminer si la réduction de la charge de phosphore s'accompagne d'une réduction correspondante de la concentration de cet élément et de la biomasse du phytoplancton. Il a été démontré, à la section précédente, que les concentrations de

phosphore et de chlorophylle a étaient à la baisse. Il est important, pour comparer les variations des teneurs en chlorophylle a et en phosphore, de mettre en corrélation des quantités ne présentant pas de variabilité spatiale et saisonnière. Ainsi, on pourrait étudier les effets annuels (les valeurs annuelles après élimination des effets saisonniers et spatiaux) obtenus à l'aide du modèle 3.4 pour les deux variables. Les effets annuels non corrigés de la chlorophylle a ont été reportés, à la figure 47, en fonction de ceux du PT. L'allure du graphique indique qu'il est raisonnable de décrire cette relation par une droite dont l'équation de régression ajustée est

$$y = 1.6 + 0.206174X$$

où X est la concentration de phosphore en milligrammes par litre et y la biomasse du phytoplancton en milligrammes par litre de chlorophylle a.

Les effets annuels moyens du phosphore, des années 1971 et 1980, sont présentés au graphique avec les valeurs estimées correspondantes de la biomasse du phytoplancton. On constate que le programme de réduction du phosphore des eaux du lac Erié a donc été efficace. A noter aussi que les teneurs en nitrate et nitrite filtrés ont présenté une tendance à la hausse pendant cette même période.

## RAPPORT ENTRE LES TENEURS EN AMMONIAC ET CELLES EN NITRITE ET NITRATE

Rockwell et coll. (1980) mentionnent qu'un rapport ammoniac/nitrite élevé indique une pollution récente tandis qu'un rapport ammoniac/nitrate faible indique un apport plus ancien suivi d'une oxydation. On présente, au tableau 42, les rapports ammoniac/nitrite et ammoniac/nitrate dans chacun des trois bassins

pour les années où des données sont disponibles. On peut y noter que le rapport est plus élevé dans le bassin central que dans les deux autres bassins et qu'il est constant, dans chacun des bassins, en 1967, 1970 et 1971. Le rapport diminue de façon très importante dans le bassin central de 0.229 en 1967 à 0.081 en 1980. Il s'agit là d'un autre indice de l'efficacité du programme de déphosphatation du lac Erié.

Tableau 5. Valeurs de EC pour les zones géographiques du lac Erié

| Date de la mission | % de variation expliquée |
|---|---|
| 14 au 23 mai 1968 | 56.2 |
| 17 au 19 juin 1968 | 81.5 |
| 31 juil. au 3 août 1968 | 75.6 |
| 1er et 2 sept. 1968 | 63.8 |
| 30 sept. au 3 oct. 1968 | 83.8 |
| 7 au 10 nov. 1968 | 34.6 |
| 31 mai au 4 juin 1969 | 42.9 |
| 3 au 7 juil. 1969 | 78.0 |
| 29 juil. au 1er août 1969 | 75.1 |
| 26 au 30 août 1969 | 87.6 |
| 15 au 18 sept. 1969 | 95.9 |
| 16 au 20 oct. 1969 | 94.8 |
| 8 au 12 déc. 1969 | 47.3 |
| 8 au 11 avril 1970 | 50.6 |
| 7 au 10 mai 1970 | 41.5 |
| 3 au 6 juin 1970 | 62.1 |
| 4 au 7 juil. 1970 | 67.1 |
| 28 juil. au 2 août 1970 | 97.2 |
| 25 au 30 août 1970 | 70.5 |
| 24 au 27 sept. 1970 | 88.2 |
| 22 au 26 oct. 1970 | 52.5 |
| 26 au 30 nov. 1970 | 35.8 |
| 14 au 18 déc. 1970 | 53.4 |
| 7 et 8 juil. 1971 | 53.4 |
| 7 au 12 fév. 1971 | 94.0 |
| 3 au 6 mars 1971 | 29.8 |
| 15 au 18 avril 1971 | 52.4 |
| 18 au 21 août 1971 | 54.53 |
| 24 au 27 nov. 1971 | 27.52 |
| 26 au 28 avril 1972 | 38.3 |
| 7 au 10 juin 1972 | 31.2 |
| 28 au 30 juin 1972 | 64.2 |
| 3 au 5 août 1972 | 95.2 |
| 29 août au 1er sept. 1972 | 82.7 |
| 29 sept. au 1er oct. 1972 | 44.3 |
| 11 au 14 nov. 1972 | 54 |
| 13 au 16 avril 1973 | 73.3 |
| 26 au 30 juil. 1973 | 77.1 |
| 29 au 31 août 1973 | 48.5 |
| 10 au 13 nov. 1973 | 72.8 |
| 24 au 27 avril 1974 | 52.63 |
| 23 au 25 août 1974 | 46.34 |
| 6 au 10 avril 1975 | 63.9 |
| 13 au 22 mai 1975 | 79.9 |
| 25 au 29 juin 1975 | 74.9 |
| 6 au 11 août 1975 | 69.0 |
| 8 au 11 oct. 1975 | 63.9 |
| 28 au 31 oct. 1975 | 50.8 |
| 26 au 30 nov. 1975 | 45.0 |
| 10 au 13 avril 1976 | 39.0 |
| 28 au 30 mai 1976 | 51.3 |

Tableau 6. Résumé des étapes de modélisation du niveau d'eau

| Modèle | Equation | E.T. | F |
|---|---|---|---|
| Initial | $Y_t$ = 570.405-0.661 sin(πt/13.75) + 0.159 cos(πt/13.75) | 0.807 | - |
| Modifié | $Y_t$ = 569.985+0.011(t-40.5) + 0.000 787(t-40.5)$^2$ - 0.568 sin(πt/13.75) + 0.167 cos(πt/13.75) | 0.673 | 17.79** |
| Définitif | $Y_t$ = 569.986+0.016(t-40.5) + 0.000 785(5-40.5)$^2$ - 0.499 sin(πt/13.75) + 0.180 cos(πt/13.75) + 0.362 sin(πt/20) - 0.022 cos(πt/20) | 0.638 | 5.23** |

** Significatif au niveau de 1 %.

Tableau 7. Température de l'air du bassin central : moyennes mensuelles (°C)

| Année | Avril | Mai | Juin | Juil. | Août | Sept. | Oct. | Nov. |
|---|---|---|---|---|---|---|---|---|
| 1967 | 5.0 | 9.1 | 18.6 | 20.1 | 20.5 | 18.1 | 13.5 | 7.2 |
| 1968 | 5.6 | 9.9 | 16.5 | 20.0 | 22.0 | 20.4 | 14.5 | 9.6 |
| 1969 | 5.6 | 10.1 | 15.9 | 21.5 | 23.1 | 20.1 | 14.3 | 7.2 |
| 1970 | 6.2 | 11.1 | 17.3 | 20.6 | 22.5 | 21.5 | 14.7 | 9.1 |
| 1971 | 3.6 | 9.5 | 18.4 | 20.8 | 21.4 | 20.9 | 15.7 | 9.7 |
| 1972 | 4.0 | 11.0 | 16.3 | 19.7 | 21.6 | 19.5 | 13.8 | 8.2 |
| 1973 | 5.4 | 10.5 | 18.1 | 21.7 | 23.3 | 20.0 | 15.4 | 8.1 |
| 1974 | 5.2 | 9.7 | 15.9 | 20.6 | 21.8 | 17.9 | 12.8 | 7.9 |
| 1975 | 3.1 | 11.6 | 17.8 | 23.0 | 22.0 | 17.8 | 14.0 | 10.1 |
| 1976 | 5.5 | 10.3 | 17.8 | 20.2 | 20.7 | 18.1 | 12.4 | 4.5 |
| 1977 | 5.4 | 11.4 | 16.7 | 21.8 | 22.3 | 20.0 | 13.2 | 8.9 |
| 1978 | 3.6 | 11.7 | 17.9 | 21.1 | 22.8 | 20.3 | 14.2 | 8.5 |
| Moyenne | 4.83 | 10.49 | 17.22 | 20.93 | 22.00 | 19.55 | 14.04 | 8.25 |
| E.T. | 0.98 | 0.86 | 1.00 | 0.94 | 0.87 | 1.27 | 0.98 | 1.51 |

Tableau 8. Résidus et sommes cumulatives des résidus pour la température (°C) de l'air du bassin central

| Année | | Avril | Mai | Juin | Juil. | Août | Sept. | Oct. | Nov. |
|---|---|---|---|---|---|---|---|---|---|
| 1967 | Résidu | 0.17 | -1.39 | 1.38 | -0.83 | -1.5 | -1.45 | -0.54 | -1.05 |
| | RC | 0.17 | -1.22 | 0.16 | -0.67 | -2.17 | -3.62 | -4.16 | -5.21 |
| 1968 | Résidu | 0.47 | -0.59 | -0.72 | -0.93 | 0.0 | 0.85 | 0.46 | 1.35 |
| | RC | 0.47 | -0.12 | -0.84 | -1.77 | -1.77 | -0.92 | -0.46 | 0.89 |
| 1969 | Résidu | 0.77 | -0.39 | -1.32 | 0.57 | 1.1 | 0.55 | 0.26 | -1.05 |
| | RC | 0.77 | 0.38 | -0.94 | -0.37 | 0.73 | 1.28 | 1.54 | 0.49 |
| 1970 | Résidu | 1.37 | 0.61 | 0.08 | -0.33 | 0.50 | 1.95 | 0.66 | 0.85 |
| | RC | 1.37 | 1.98 | 2.06 | 1.73 | 2.23 | 4.18 | 4.84 | 5.69 |
| 1971 | Résidu | -1.23 | -0.99 | 1.18 | -0.13 | +0.60 | 1.35 | 1.66 | 1.45 |
| | RC | -1.23 | -2.22 | -1.04 | -1.17 | -1.77 | -0.42 | 1.24 | 2.69 |
| 1972 | Résidu | -0.88 | 0.51 | -0.92 | -.123 | -0.40 | -0.05 | -0.24 | -0.05 |
| | RC | -0.88 | -0.37 | -1.29 | -2.52 | -2.92 | -2.97 | -3.21 | -3.26 |
| 1973 | Résidu | 0.57 | 0.01 | 0.88 | 0.77 | 1.30 | 0.45 | 1.36 | -0.15 |
| | RC | 0.57 | 0.58 | 1.46 | 2.23 | 3.53 | 3.98 | 5.34 | 5.19 |
| 1974 | Résidu | 0.37 | -0.79 | -1.32 | -0.33 | -0.20 | -1.65 | -1.24 | -0.35 |
| | RC | 0.37 | -0.42 | -1.74 | -2.07 | -2.27 | -3.92 | -5.16 | -5.51 |
| 1975 | Résidu | -1.73 | 1.11 | 0.58 | -2.07 | 0.0 | -1.75 | -0.04 | 1.85 |
| | RC | -1.73 | -0.62 | -0.04 | 2.03 | 2.03 | 0.28 | 0.24 | 2.09 |
| 1976 | Résidu | 0.67 | -0.19 | 0.58 | -0.73 | -1.30 | -1.45 | -1.64 | -3.75 |
| | RC | 0.67 | 0.48 | 1.06 | 0.33 | -0.97 | -2.42 | -4.06 | -7.13 |
| 1977 | Résidu | 0.57 | 0.91 | -0.52 | 0.87 | 0.30 | 0.45 | -0.84 | 0.65 |
| | RC | 0.57 | 1.48 | 0.96 | 1.83 | 2.13 | 2.58 | 1.74 | 2.39 |
| 1978 | Résidu | -1.23 | 1.21 | 0.68 | 0.17 | 0.80 | 0.75 | 0.16 | 0.25 |
| | RC | -1.23 | -0.02 | 0.66 | 0.83 | 1.63 | 2.38 | 2.54 | 2.79 |

Tableau 9. Résumé de la structure de la température de l'air du bassin central, 1967 à 1978

| Année | Printemps | Eté | Automne |
|---|---|---|---|
| 1967 | N | F | F |
| 1968 | N | F | C |
| 1969 | N | N | N |
| 1970 | C | C | C |
| 1971 | F | N | C |
| 1972 | F | F | F |
| 1973 | C | C | C |
| 1974 | N | F | F |
| 1975 | N | C | N |
| 1976 | N | F | F |
| 1977 | F | N | N |
| 1978 | N | C | N |

C - chaud.
F - froid.
N - normal.

Tableau 10. Résultats de l'ajustement du modèle 3.4 aux températures de surface du lac Erié - 1967 à 1978

| Année | Bassin ouest | | | | | Bassin central | | | | | Bassin est | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Température maximale $\alpha + R_t$ | Phase $\phi$ | Nombre de jours $D_t$ | $R^2$ | F | Température maximale $\alpha + R_t$ | Phase $\phi$ | Nombre de jours $D_t$ | $R^2$ | F | Température maximale $\alpha + R_t$ | Phase $\phi$ | Nombre de jours $D_t$ | $R^2$ | F |
| 1967 | 22.87 | -1.768 | 205.4 | 0.92 | 37.9** | 22.25 | -1.884 | 206.8 | 0.76 | 11.3** | 22.82 | -1.929 | 224.0 | 0.89 | 24.3** |
| 1968 | 24.14 | -1.877 | 218.0 | 0.98 | 60.7** | 22.52 | -1.970 | 228.8 | 0.99 | 131.4** | 22.07 | -2.005 | 232.8 | 0.99 | 109.6** |
| 1969 | 23.43 | -1.907 | 221.4 | 0.99 | 141.3** | 22.39 | -1.971 | 229.0 | 0.98 | 85.5** | 19.78 | -1.851 | 215.0 | 0.86 | 14.9** |
| 1970 | 23.99 | -1.873 | 217.6 | 0.97 | 130.0** | 22.87 | -1.956 | 227.2 | 0.96 | 78.2** | 22.14 | -1.989 | 231.0 | 0.93 | 45.7** |
| 1971 | 21.47 | -1.749 | 203.2 | 0.86 | 9.0** | 19.46 | -1.882 | 218.6 | 0.81 | 6.4** | 21.02 | -1.802 | 209.2 | 0.95 | 18.2** |
| 1972 | 22.53 | -1.880 | 218.4 | 0.94 | 32.3** | 21.65 | -1.963 | 228.0 | 0.96 | 50.7** | 21.31 | -1.994 | 231.6 | 0.97 | 62.9** |
| 1973 | 25.60 | -1.809 | 210.0 | 0.99 | 76.4** | 23.90 | -1.879 | 218.2 | 0.99 | 91.6** | 23.78 | -1.882 | 218.6 | 0.99 | 809.5** |
| 1974 | Aucune donnée | | | | | | | | | | | | | | |
| 1975 | 23.09 | -1.866 | 216.8 | 0.96 | 47.6** | 21.26 | -1.937 | 225.0 | 0.94 | 41.6** | 21.62 | -1.938 | 225.0 | 0.93 | 26.1** |
| 1977 | -- | -- | -- | | | 23.17 | -1.935 | 224.8 | 0.99 | 280.2** | 22.34 | -1.970 | 228.8 | 0.99 | 211.4** |
| 1978 | -- | -- | -- | | | 22.65 | -1.968 | 228.6 | 0.97 | 129.8 | 22.67 | -1.965 | 228.2 | 0.98 | 220.2** |

** Significatif au niveau de 1 %.

Tableau 11. Valeurs de la statistique F pour le test d'égalité des cycles annuels de température de l'eau

| Bassin ouest | | Bassin central | | Bassin est | |
|---|---|---|---|---|---|
| Années | Statistique F | Années | Statistique F | Années | Statistique F |
| 1967-1975 | 2.77* | 1967-1978 | 3.65** | 1967-1978 | 5.70** |

* Significatif au niveau de 5 %.
** Significatif au niveau de 1 %.

Tableau 12. Température moyenne de l'hypolimnion, température au jour 200 du calendrier et pente et ordonnée à l'origine de la droite de régression, 1967 à 1980

| Année | Moyenne | Température (Jour 100 du calendrier) | Ordonnée à l'origine | Pente |
|---|---|---|---|---|
| 1967 | 10.92 | 10.28 | 10.59 | 0.0021 |
| 1968 | 13.18 | 13.31 | 13.51 | -0.0042 |
| 1969 | 11.36 | 10.96 | 8.51 | 0.0489 |
| 1970 | 9.90 | 9.99 | 7.31 | 0.0536 |
| 1971 | 10.96 | 10.70 | 9.40 | 0.0260 |
| 1972 | 10.98 | 10.36 | 6.64 | 0.0743 |
| 1973 | 11.39 | 10.78 | 10.36 | 0.0083 |
| 1975 | 8.17 | 8.86 | 5.48 | 0.0676 |
| 1977 | 11.72 | 11.78 | 8.42 | 0.0672 |
| 1978 | 8.57 | 8.87 | 7.31 | 0.0313 |
| 1979 | 10.15 | 9.48 | 7.68 | 0.0359 |
| 1980 | 11.15 | 9.31 | 5.09 | 0.0843 |

Tableau 13. Valeurs estimées des paramètres du modèle 3.4 pour la chlorophylle a (bassin ouest)

| Années | Paramètres du modèle $\hat{\alpha}_t$ | $\hat{\beta}_t$ | $\hat{\gamma}_t$ | $\hat{\sigma}$ | Maximum estimé | Coefficient de détermination |
|---|---|---|---|---|---|---|
| 1968 | 9.00 | -0.708 | -1.364 | 2.283 | 10.54 | 0.82 |
| 1970 | 9.77 | -1.238 | 0.736 | 4.606 | 11.21 | 0.74 |
| 1972 | 9.75 | 3.038 | 1.935 | 1.924 | 13.35 | 0.77 |

Tableau 14. Valeurs de $F_1$, $F_2$, et $F_0$ pour la chlorophylle a (bassin ouest)

| $H_1$ | | $H_2$ | | $H_0$ | |
|---|---|---|---|---|---|
| Degrés de liberté | $F_1$ | Degrés de liberté | $F_2$ | Degrés de liberté | $F_0$ |
| 4.9 | 1.61 | 2.13 | 1.55 | 6.9 | 1.30 |

Tableau 15. Valeurs estimées des paramètres du modèle 3.6 pour la chlorophylle a (bassin ouest)

| $\hat{\alpha}_t$ | | | | |
|---|---|---|---|---|
| 1968 | 1970 | 1972 | $\hat{\beta}$ | $\hat{\gamma}$ |
| 8.27 | 10.01 | 11.81 | -0.1585 | 0.2579 |

Tableau 16. Valeurs estimées des paramètres du modèle 3.4 pour la chlorophylle a (bassin ouest)

| | Paramètres du modèle | | | | Maximum estimé | Coefficient de détermination |
|---|---|---|---|---|---|---|
| Année | $\hat{\alpha}_0$ | $\hat{\alpha}_1$ | $\hat{\alpha}_a$ | $\hat{\sigma}$ | | |
| 1968 | 5.163 | -0.751 | -1.478 | 1.299 | 6.82 | 0.61 |
| 1970 | 6.425 | -0.369 | -0.276 | 2.647 | 6.89 | 0.89 |
| 1979 | 3.998 | 0.666 | -0.275 | 2.589 | 4.718 | 0.69 |
| 1980 | 3.268 | -0.355 | -0.739 | 0.637 | 4.087 | 0.65 |

Tableau 17. Valeurs de $F_1$, $F_2$, et $F_0$ pour la chlorophylle a (bassin central)

| $H_1$ | | $H_2$ | | $H_0$ | |
|---|---|---|---|---|---|
| Degrés de liberté | $F_1$ | Degrés de liberté | $F_2$ | Degrés de liberté | $F_0$ |
| 6.23 | 0.266 | 3.29 | 4.304* | 9.23 | 1.39 |

*Significatif au niveau de 5 %.

Tableau 18. Valeurs estimées des paramètres du modèle 3.6 pour la chlorophylle a (bassin central)

| $\hat{a}_t$ | | | | $\hat{\beta}$ | $\hat{\gamma}$ |
|---|---|---|---|---|---|
| 1968 | 1970 | 1979 | 1980 | | |
| 5.045 | 6.345 | 4.327 | 3.045 | 0.766 | -0.5909 |

Tableau 19. Valeurs estimées des paramètres du modèle 3.4 pour la chlorophylle a (bassin est)

| Année | Paramètres du modèle | | | | Maximum estimé | Coefficient de détermination |
|---|---|---|---|---|---|---|
| | $\hat{a}_0$ | $\hat{a}_1$ | $\hat{a}_a$ | $\hat{\sigma}$ | | |
| 1968 | 7.962 | 4.538 | -1.952 | 2.08 | 8.241 | 0.55 |
| 1970 | 5.334 | -0.031 | 0.505 | 1.65 | 5.808 | 0.62 |
| 1972 | 3.735 | 0.153 | -0.424 | 0.885 | 4.186 | 0.63 |
| 1980 | 2.698 | 0.575 | 0.061 | 0.652 | 3.276 | 0.68 |

Tableau 20. Valeurs de $F_1$, $F_2$, et $F_0$ pour la chlorophylle a (bassin est)

| $H_1$ | | $H_2$ | | $H_0$ | |
|---|---|---|---|---|---|
| Degrés de liberté | $F_1$ | Degrés de liberté | $F_2$ | Degrés de liberté | $F_0$ |
| 6.16 | 0.949 | 3.22 | 6.175** | 9.16 | 2.492 |

**Significatif au niveau de 1 %.

Tableau 21. Valeurs estimées des paramètres du modèle 3.6 pour la chlorophylle a (bassin est)

| $\hat{a}_t$ | | | | $\hat{\beta}$ | $\hat{\gamma}$ |
|---|---|---|---|---|---|
| 1968 | 1970 | 1979 | 1980 | | |
| 5.378 | 5.233 | 3.675 | 2.850 | 0.3582 | -0.0670 |

Tableau 22. Valeurs estimées des paramètres du modèle 3.4 pour le PT (bassin ouest)

| Année | Paramètres du modèle $\hat{\alpha}$ | $\hat{\beta}$ | $\hat{\gamma}$ | $\hat{\sigma}$ | Maximum estimé |
|---|---|---|---|---|---|
| 1968 | 0.009 22 | 0.029 98 | 0.028 277 | 0.002 55 | 0.0375 |
| 1970 | 0.043 99 | -0.003 39 | 0.004 68 | 0.018 55 | 0.0393 |
| 1971 | 0.045 21 | 0.015 3 | 0.024 78 | 0.022 102 | 0.029 |
| 1972 | 0.031 37 | -0.002 64 | 0.001 25 | 0.003 49 | 0.028 |
| 1977 | 0.018 7 | 0.002 43 | 0.003 25 | 0.002 66 | 0.014 |
| 1978 | 0.016 8 | -0.000 267 | 0.002 44 | 0.004 82 | 0.014 |

Tableau 23. Valeurs de $F_1$, $F_2$, et $F_0$ pour le PT (bassin ouest)

| $H_1$ | | $H_2$ | | $H_1$ $H_2$ | |
|---|---|---|---|---|---|
| Degrés de liberté | $F_1$ | Degrés de liberté | $F_2$ | Degrés de liberté | $F_0$ |
| 10,20 | 0.719 | 5.30 | 5.475** | 15.20 | 1.778 |

**Significatif au niveau de 1 %.

Tableau 24. Valeurs estimées des paramètres du modèle 3.6 pour le PT (bassin ouest)

| $\hat{a}_t$ | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1968 | 1970 | 1971 | 1972 | 1977 | 1978 | $\hat{\beta}$ | $\hat{\gamma}$ |
| 0.037 | 0.045 | 0.043 | 0.028 | 0.0189 | 0.0193 | 0.0193 | 0.0024 |

Tableau 25. Valeurs estimées des paramètres du modèle 3.4 pour le PT (bassin central)

| Année | Paramètres du modèle $\hat{\alpha}$ | $\hat{\beta}$ | $\hat{\gamma}$ | $\hat{\sigma}$ | Minimum |
|---|---|---|---|---|---|
| 1968 | 0.030 6 | 0.009 5 | -0.008 6 | 0.002 42 | 0.018 |
| 1970 | 0.023 1 | 0.000 48 | 0.001 5 | 0.009 75 | 0.022 |
| 1971 | 0.020 9 | -0.002 1 | 0.003 75 | 0.005 762 | 0.017 |
| 1972 | 0.019 6 | -0.003 7 | -0.001 98 | 0.004 6 | 0.015 |
| 1977 | 0.013 67 | 0.004 6 | 0.000 12 | 0.001 69 | 0.009 |
| 1978 | 0.011 89 | -0.001 01 | 0.001 1 | 0.000 77 | 0.010 |
| 1979 | 0.011 87 | 0.001 54 | -0.000 39 | 0.004 43 | 0.010 |
| 1980 | 0.010 6 | 0.001 53 | -0.000 53 | 0.001 57 | 0.009 |

Tableau 26. Valeurs de $F_1$, $F_2$, et $F_0$ pour le PT (bassin central)

| $H_1$ | | $H_2$ | | $H_0$ | |
|---|---|---|---|---|---|
| Degrés de liberté | $F_1$ | Degrés de liberté | $F_2$ | Degrés de liberté | $F_0$ |
| 14.24 | 1.81 | 1.81 | 7.233** | 21.24 | 0.354 |

**Significatif au niveau de 1 %.

Tableau 27. Valeurs estimées des paramètres du modèle 3.6 pour le PT (bassin central)

| $\hat{\alpha}_t$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1968 | 1970 | 1971 | 1972 | 1977 | 1978 | 1979 | 1980 | $\hat{\beta}$ | $\hat{\gamma}$ |
| 0.024 | 0.023 | 0.020 | 0.018 | 0.014 | 0.012 | 0.012 | 0.011 | 0.000 424 | 0.000 91 |

Tableau 28. Valeurs estimées des paramètres du modèle 3.4 pour le PT (bassin est)

| Année | Paramètres du modèle $\hat{\alpha}$ | $\hat{\beta}$ | $\hat{\gamma}$ | $\hat{\sigma}$ | Minimum |
|---|---|---|---|---|---|
| 1968 | 0.019 | 0.009 71 | -0.004 22 | 0.002 37 | 0.009 |
| 1970 | 0.018 | -0.000 96 | 0.000 41 | 0.006 49 | 0.017 |
| 1971 | 0.018 | 0.005 15 | 0.002 31 | 0.006 03 | 0.012 |
| 1972 | 0.013 | 0 002 01 | 0.002 266 | 0.006 90 | 0.010 |
| 1977 | 0.013 | -0.001 07 | -0.001 3 | 0.001 19 | 0.011 |
| 1978 | 0.012 | $-1.283 \times 10^{-6}$ | -0.001 27 | 0.001 57 | 0.011 |
| 1980 | 0.012 | -0.000 564 | 0.001 29 | 0.002 48 | 0.011 |

Tableau 29. Valeurs de $F_1$, $F_2$, et $F_0$ pour le PT (bassin est)

| $H_1$ | | $H_2$ | | $H_0$ | |
|---|---|---|---|---|---|
| Degrés de liberté | $F_1$ | Degrés de liberté | $F_2$ | Degrés de liberté | $F_0$ |
| 18,25 | 0.910 | 6,37 | 2.471* | 12,25 | 0.467 |

*Significatif au niveau de 5 %.

Tableau 30. Valeurs estimées des paramètres du modèle 3.6 pour le PT (bassin est)

| $\hat{a}_t$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1968 | 1970 | 1971 | 1972 | 1977 | 1978 | 1980 | $\hat{\beta}$ | $\hat{\gamma}$ |
| 0.0118 | 0.0180 | 0.0171 | 0.0130 | 0.0137 | 0.0125 | 0.0119 | -0.000 52 | 0.000 32 |

Tableau 31. Valeurs de $F_1$, $F_2$ et $F_0$ pour le PRS (les trois bassins)

| | $H_1$ | | $H_2$ | | $H_0$ | |
|---|---|---|---|---|---|---|
| Bassin | d.l. | $F_1$ | d.l. | $F_2$ | d.l. | $F_0$ |
| Ouest | 8,17 | 2.08 | 4,25 | 6.774** | 12,17 | 4.431** |
| Central | 12,30 | 2.26* | 6,42 | 3.71** | 18,30 | 3.19** |
| Est | 8,16 | 2.23 | 4,24 | 3.37* | 12,16 | 3.07* |

*Significatif au niveau de 5 %.
**Significatif au niveau de 1 %.
d.l.-Degrés de liberté.

Tableau 32. Valeurs estimées des paramètres du modèle 3.6 pour le PRS (les trois bassins)

| | $\hat{a}_t$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Bassin | 1968 | 1970 | 1971 | 1973 | 1978 | 1979 | 1980 | $\hat{\beta}$ | $\hat{\gamma}$ |
| Ouest | 0.0121 | 0.007 | 0.002 9 | 0.0079 | 0.002 15 | --- | --- | 0.000 74 | -0.000 046 |
| Central | 0.0059 | 0.0036 | 0.001 69 | 0.0049 | 0.001 74 | 0.001 47 | 0.000 522 | 0.000 76 | 0.000 51 |
| Est | 0.0026 | 0.0030 | 0.002 9 | 0.0083 | --- | 0.001 4 | | -0.000 71 | 0.000 08 |

Tableau 33. Valeurs estimées de $a_t$ et de $R_t$ pour le $N\text{-}NO_3NO_2$ (les trois bassins)

| | Bassin ouest | | Bassin central | | Bassin est | |
|---|---|---|---|---|---|---|
| Année | $\hat{a}_t$ | $R_t$ | $\hat{a}_t$ | $R_t$ | $\hat{a}_t$ | $R_t$ |
| 1967 | 0.18 | 0.13 | 0.114 | 0.092 | 0.079 | 0.081 |
| 1968 | 0.18 | 0.15 | 0.120 | 0.107 | 0.100 | 0.106 |
| 1970 | 0.23 | 0.18 | 0.117 | 0.099 | 0.119 | 0.105 |
| 1971 | 0.14 | 0.14 | 0.082 | 0.106 | 0.091 | 0.100 |
| 1978 | 0.25 | 0.21 | 0.206 | 0.100 | 0.180 | 0.094 |
| 1980 | -- | -- | 0.212 | 0.100 | -- | -- |

Tableau 34. Valeurs de $F_1$, $F_2$ et $F_0$ pour le $N\text{-}NO_3NO_2$ (les trois bassins)

| | $H_2$ | | $H_0$ | |
|---|---|---|---|---|
| Bassin | d.l. | $F_2$ | d.l. | $F_0$ |
| Ouest | 8,20 | 9.13** | 12,20 | 17.26** |
| Central | 10,26 | 4.29** | 15,26 | 8.37** |
| Est | 9,19 | 3.89** | 12,19 | 6.77** |

**Significatif au niveau de 1 %.
d.l.-Degrés de liberté.

Tableau 35. Valeurs estimées de $a_t$ et de $R_t$ pour le $N\text{-}NO_3NO_2$ (les trois bassins)

| Année | Bassin ouest | | Bassin central | | Bassin est | |
|---|---|---|---|---|---|---|
| | $\hat{a}_t$ | $R_t$ | $\hat{a}_t$ | $R_t$ | $\hat{a}_t$ | $R_t$ |
| 1967, 1968 et 1970 | 0.217 | 0.178 | 0.118 | 0.099 | 0.111 | 0.105 |
| 1971 | 0.138 | 0.140 | 0.082 | 0.106 | 0.091 | 0.100 |
| 1978 et 1980 | --- | --- | 0.209 | 0.101 | --- | --- |

Tableau 36. Valeurs de $F_1$, $F_2$ et $F_0$ pour le PRS (les trois bassins)

| Bassin | $H_1$ | | $H_2$ | | $H_0$ | |
|---|---|---|---|---|---|---|
| | d.l. | $F_1$ | d.l. | $F_2$ | d.l. | $F_0$ |
| Ouest | 6,16 | 0.48 | 3,22 | 3.07* | 9,16 | 1.20** |
| Central | 8,22 | 3.00* | 4,30 | 57.23** | 12,22 | 31.26** |
| Est | 6,15 | 2.17 | 3,21 | 5.79** | 9,15 | 4.02** |

*Significatif au niveau de 5 %.
**Significatif au niveau de 1 %.
d.l.-Degrés de liberté.

Tableau 37. Valeurs estimées des paramètres du modèle 3.6 pour le PRS (les trois bassins)

| Bassin | $\hat{a}_t$ | | | | | $\hat{\beta}_t$ | $\hat{\gamma}_t$ |
|---|---|---|---|---|---|---|---|
| | 1968 | 1970 | 1971 | 1978 | 1980 | | |
| Ouest | 0.023 | 0.027 | 0.017 | 0.008 | --- | 0.005 | 0.001 |
| Central | 0.027 | 0.030 | 0.020 | 0.011 | 0.017 | 0.007 | 0.005 |
| Est | 0.023 | 0.027 | 0.027 | 0.008 | --- | -0.0053 | 0.0012 |

Tableau 38. Résultats de l'ajustement du modèle 3.4 aux mesures de la profondeur au disque de Secchi - lac Erié, 1967 à 1978

| Année | Bassin ouest | | | | | Bassin central | | | | | Bassin est | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Amplitude $\alpha + R$ | Phase $\phi$ | Nombre de jours $D_t$ | $R^2$ | F | Amplitude $\alpha + R$ | Phase $\phi$ | Nombre de jours $D_t$ | $R^2$ | F | Amplitude $\alpha + R$ | Phase $\phi$ | Nombre de jours $D_t$ | $R^2$ | F |
| 1967 | 3.27 | 0.32 | 36.75 | 0.71 | 8.44* | 5.49 | 1.77 | 206.1 | 0.86 | 22.05** | 7.84 | 1.89 | 219.51 | 0.91 | 23.34** |
| 1968 | 1.82 | 1.68 | 195.59 | 0.50 | 0.986 | 3.17 | 1.76 | 204.90 | 0.76 | 4.86* | 4.20 | 2.03 | 235.38 | 0.82 | 4.49** |
| 1969 | 2.25 | 1.83 | 213.15 | 0.42 | 1.09 | 3.40 | 1.80 | 208.71 | 0.43 | 1.48 | 4.18 | 1.70 | 197.64 | 0.48 | 2.28* |
| 1970 | 1.68 | 1.62 | 188.09 | 0.49 | 3.30* | 4.53 | 1.84 | 213.21 | 0.86 | 22.29** | 4.95 | 1.88 | 218.07 | 0.79 | 13.07** |
| 1971 | -- | 3.91 | -- | 0.20 | 0.37 | 4.20 | 1.61 | 187.20 | 0.55 | 1.80 | 5.72 | 1.58 | 183.59 | 0.80 | 3.90** |
| 1972 | 2,57 | 1.76 | 204.32 | 0.45 | 0.82 | 4.19 | 1.72 | 199.27 | 0.67 | 4.05* | 3.98 | 2.01 | 233.41 | 0.65 | 2.83* |
| 1973 | | | | | | 3.74 | 1.85 | 214.73 | 0.99 | 193.2** | 4.27 | 1.81 | 209.86 | 0.82 | 4.53* |
| 1975 | 2.03 | 1.77 | 205.36 | 0.40 | 1.34 | 5.14 | 1.82 | 210.97 | 0.87 | 16.77* | 4.89 | 1.88 | 218.7 | 0.85 | 11.56** |
| 1977 | | | | | | 5.00 | 1.92 | 222.25 | 0.63 | 4.20** | 4.87 | 2.07 | 240.22 | 0.76 | 7.86** |
| 1978 | | | | | | 5.62 | 1.50 | 173.69 | 0.33 | 1.99* | 5.13 | 1.80 | 208.59 | 0.23 | 1.04 |
| Toutes les années | 1.89 | 1.67 | 193.59 | 0.27 | 9.63** | 4.37 | 1.68 | 195.49 | 0.48 | 34.17** | 4.61 | 1.78 | 207.35 | 0.39 | 20.85** |

* Significatif au niveau de 5 %
** Significatif au niveau de 1 %

Tableau 39. Valeurs de la statistique F pour le test d'égalité d'une année à l'autre des cycles des profondeurs au disque de Secchi

| Bassin ouest | | Bassin central | | Bassin est | |
|---|---|---|---|---|---|
| Années | Statistique F | Années | Statistique F | Années | Statistique F |
| 1967-1975 | 0.984 | 1967-1978 | 2.384** | 1967-1978 | 3.475** |

** Significatif au niveau de 1 %.

Tableau 40. Résultats de l'ajustement du modèle 3.4 aux données sur les mesures de turbidité - lac Erié, 1967 à 1972

| Année | Bassin ouest | | | | | Bassin central | | | | | Bassin est | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Amplitude | | Phase | Nombre de jours | | Amplitude | | Phase | Nombre de jours | | Amplitude | | Phase | Nombre de jours | |
| | $\alpha_{it}$ + | $R_{it}$ | $\phi_{it}$ | $D_{it}$ | $R^2$ | F | $\alpha_{it}$ + | $R_{it}$ | $\phi_{it}$ | $D_{it}$ | $R^2$ | F | $\alpha_{it}$ + | $R_{it}$ | $\phi_{it}$ | $D_{it}$ | $R^2$ | F |
| 1967 | -- | | -- | -- | -- | -- | -0.095 | 1.36 | 206.43 | 0.78 | 12.72** | -0.51 | 1.27 | 217.11 | 0.94 | 40.55** | | |
| 1968 | -0.07 | 1.35 | 208.20 | 0.98 | 83.92** | -1.68 | 1.25 | 219.72 | 0.81 | 6.4* | -1.66 | 1.08 | 239.74 | 0.94 | 14.87** | | |
| 1969 | -1.58 | 1.39 | 199.64 | 0.917 | 22.18** | -1.58 | 1.97 | 136.65 | 0.71 | 5.02* | -1.59 | 1.47 | 194.94 | 0.94 | 40.07** | | |
| 1970 | 1.00 | 1.35 | 208.03 | 0.699 | 6.97* | -0.38 | 1.36 | 207.09 | 0.86 | 18.18** | -0.566 | 1.52 | 188.2 | 0.74 | 8.55** | | |
| 1971 | 0.77 | 1.56 | 183.61 | 0.98 | 44.7** | -0.59 | 1.43 | 198.32 | 0.96 | 23.59** | -0.92 | 1.43 | 198.67 | 0.99 | 2627.7** | | |
| 1972 | 0.89 | 1.31 | 212.5 | 0.62 | 3.23** | -0.19 | 1.29 | 215.70 | 0.45 | 1.64 | -0.15 | 1.01 | 247.1 | 0.21 | 0.54** | | |
| Toutes les années | | 1.47 | 194.49 | 0.59 | 28.12** | -0.56 | 1.37 | 205.75 | 0.53 | 23.56** | -0.69 | 1.37 | 206.15 | 0.50 | 18.68** | | |

* Significatif au niveau de 5 %.
** Significatif au niveau de 1 %.

Tableau 41. Valeurs de la statistique F pour le test d'égalité d'une année à l'autre des cycles saisonniers de la turbidité

| Bassin ouest | | Bassin central | | Bassin est | |
|---|---|---|---|---|---|
| Années | Statistique F | Années | Statistique F | Années | Statistique F |
| 1967-1972 | 2.970** | 1967-1972 | 8.956** | 1967-1972 | 2.905** |

** Significatif au niveau de 1 %.

Tableau 42. Rapport de l'ammoniac au nitrite et au nitrate

| Année | Bassin ouest | Bassin central | Bassin est |
|---|---|---|---|
| 1967 | 0.106 | 0.229 | 0.207 |
| 1970 | 0.124 | 0.254 | 0.243 |
| 1971 | 0.123 | 0.244 | - |
| 1978 | - | 0.053 | - |
| 1980 | - | 0.081 | - |

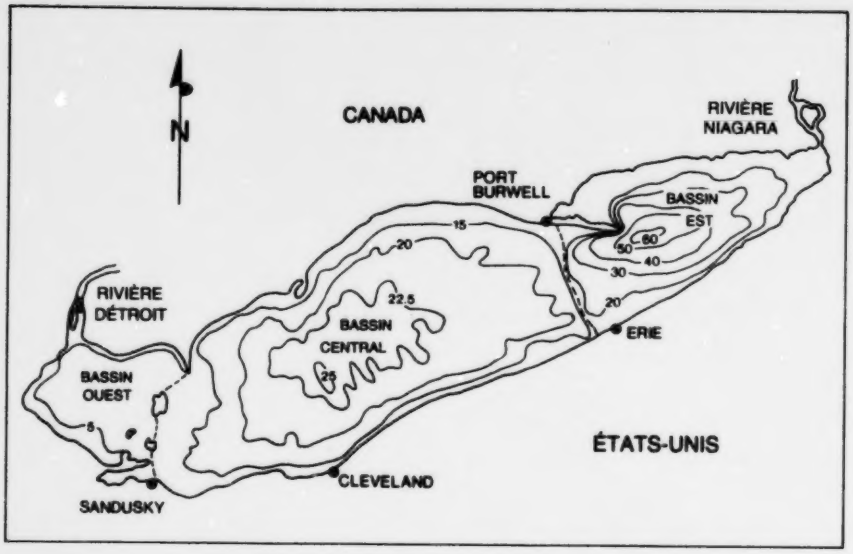



**Figure 6. Limites des bassins ouest, central et est du lac Érié.**

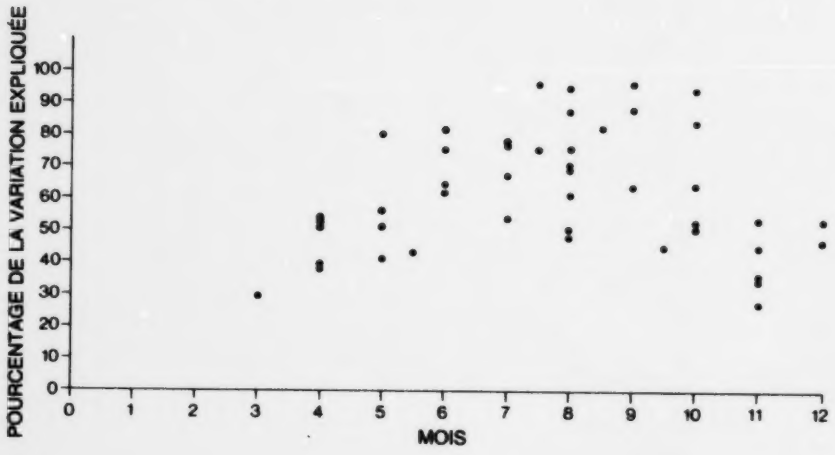



**Figure 7. Valeurs de EC en fonction du mois et de la température (1968-1976).**

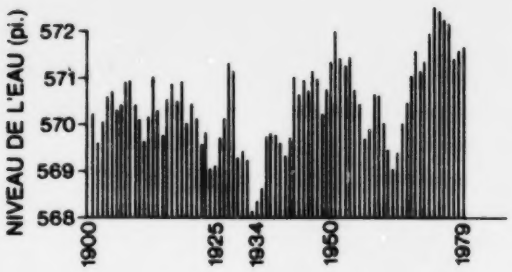



**Figure 8. Niveau annuel moyen de l'eau du bassin central.**

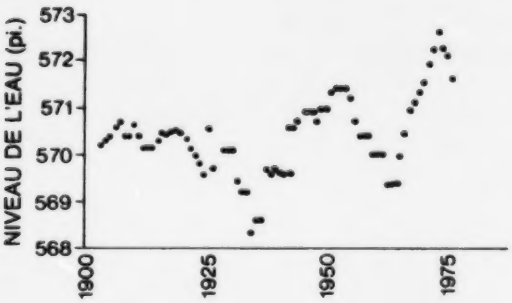



**Figure 9. Médiane mobile de cinq ans.**

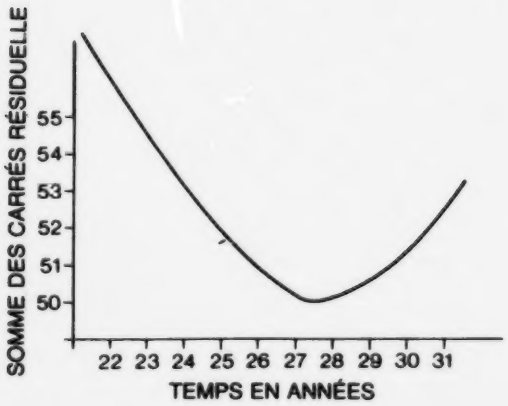



**Figure 10. Somme des carrés résiduelle en fonction de l'année.**

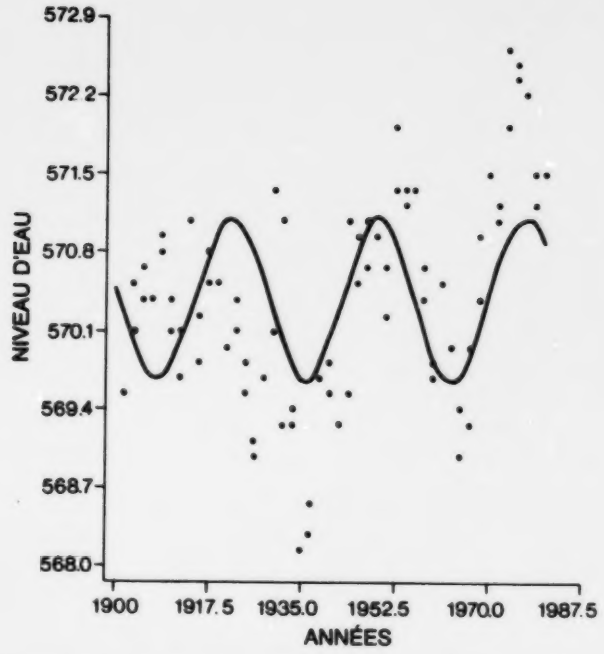



**Figure 11.** Niveaux d'eau observés et estimés en fonction de l'année.

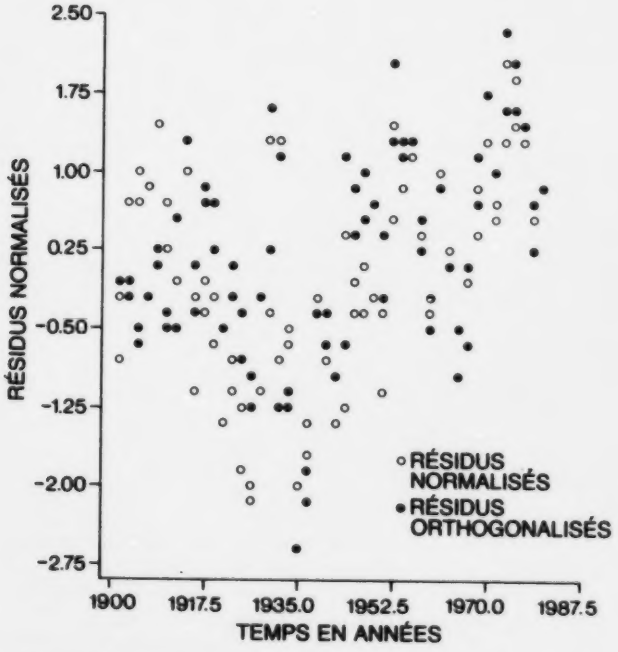



**Figure 12.** Résidus normalisés et orthogonalisés en fonction de l'année.

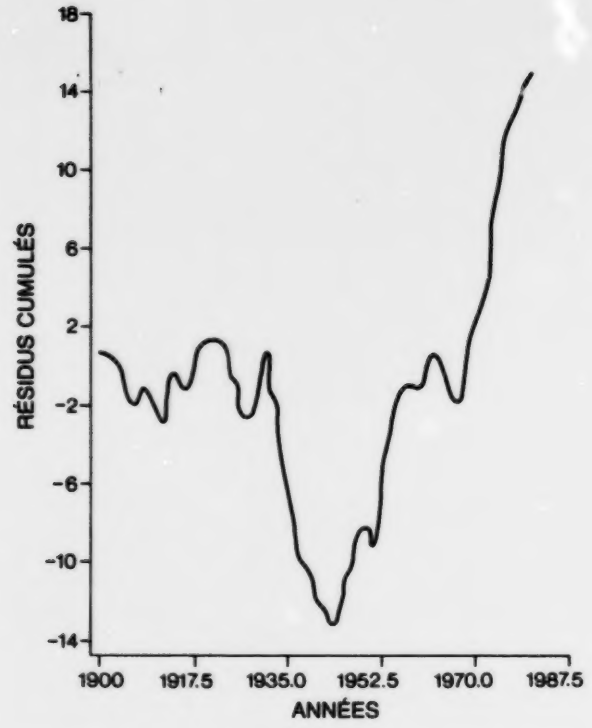



**Figure 13. Résidus normalisés cumulés en fonction de l'année.**

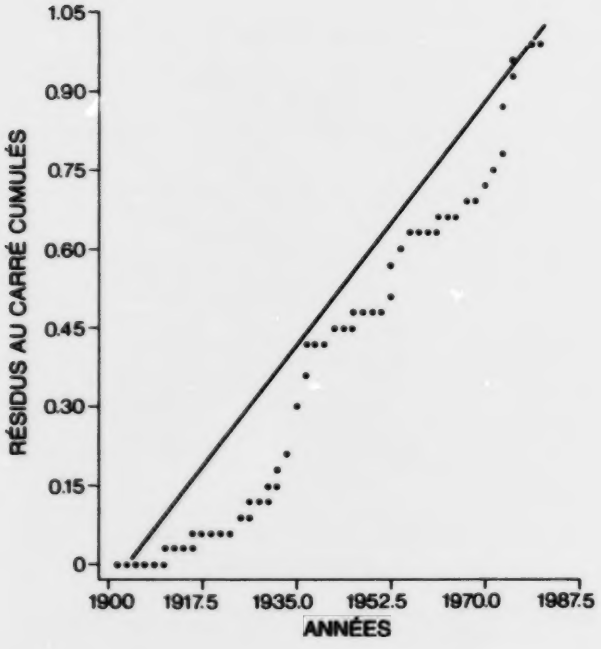



**Figure 14. Résidus au carré cumulés et leurs valeurs prévues.**

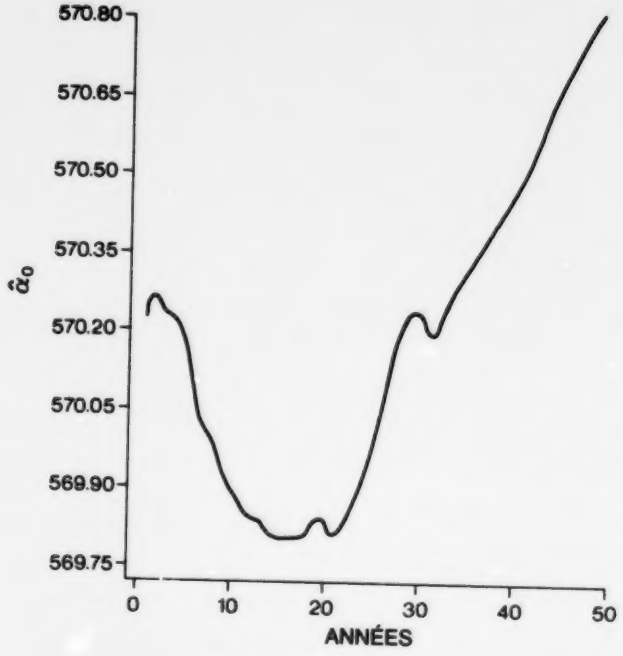



**Figure 15. Valeurs de $\hat{\alpha}_0$ en fonction de l'année.**

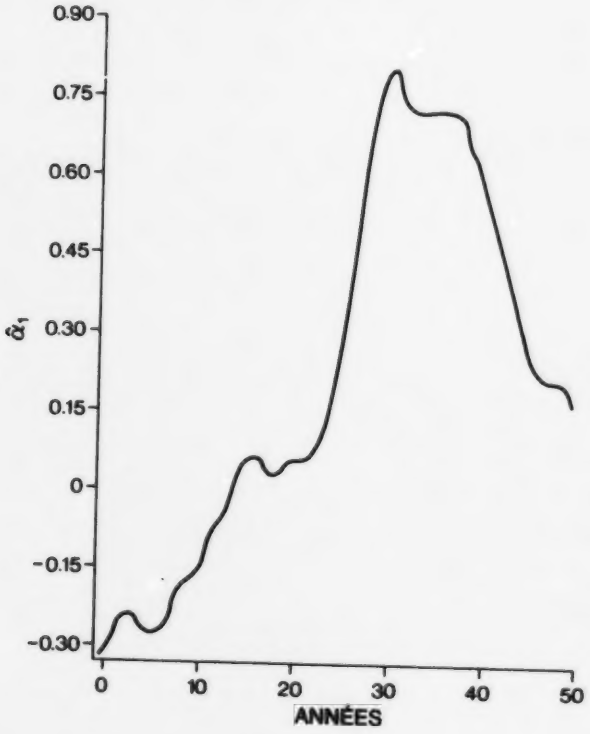



**Figure 16. Valeurs de $\hat{\alpha}_1$ en fonction de l'année.**

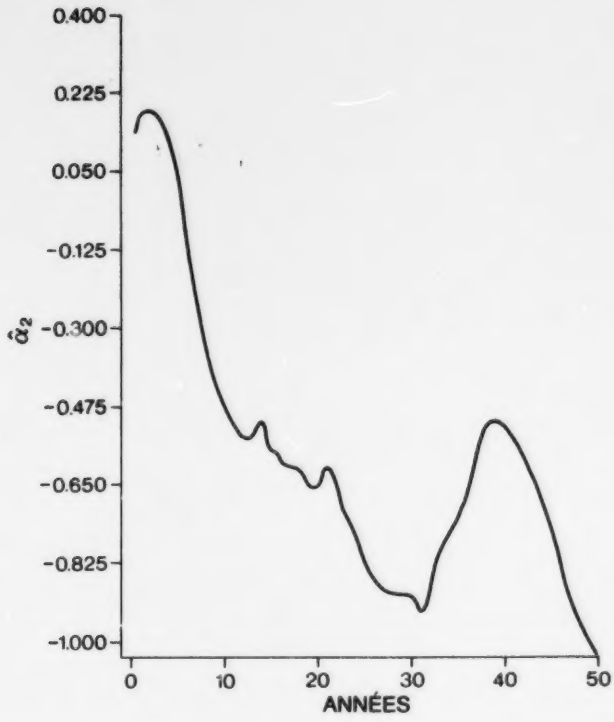



**Figure 17. Valeurs de $\hat{\alpha}_1$ en fonction de l'année.**

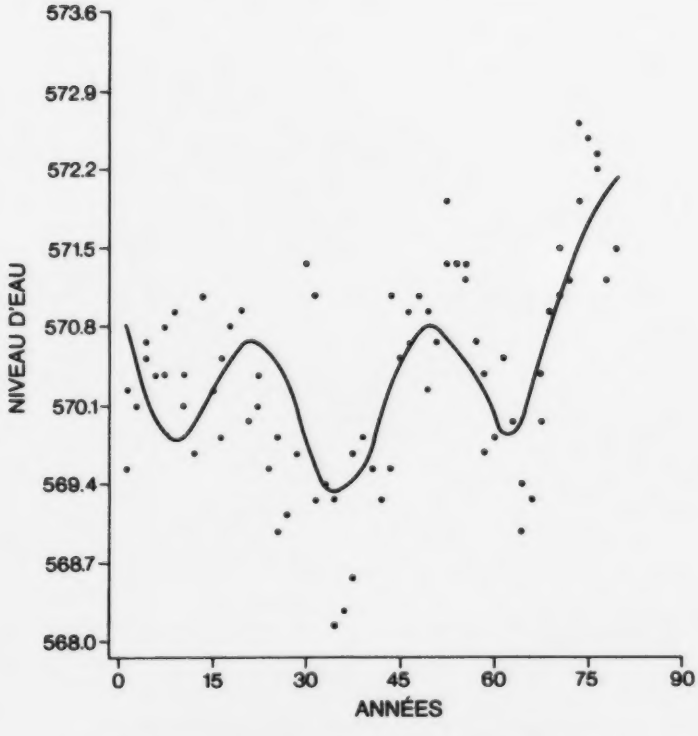



**Figure 18. Niveaux d'eau observés et estimés à l'aide du modèle 3.3.**

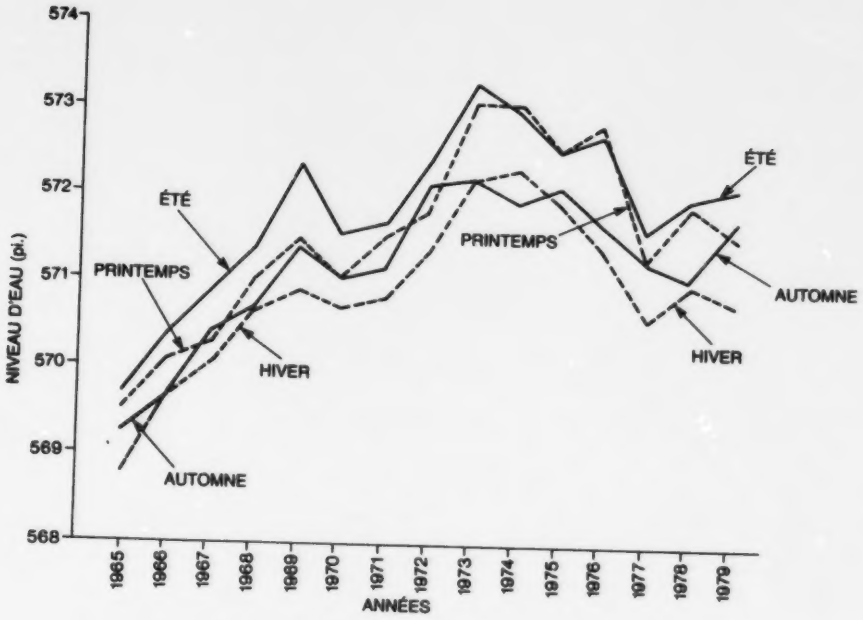



**Figure 19. Niveau d'eau moyen au printemps, en été, en automne et en hiver.**

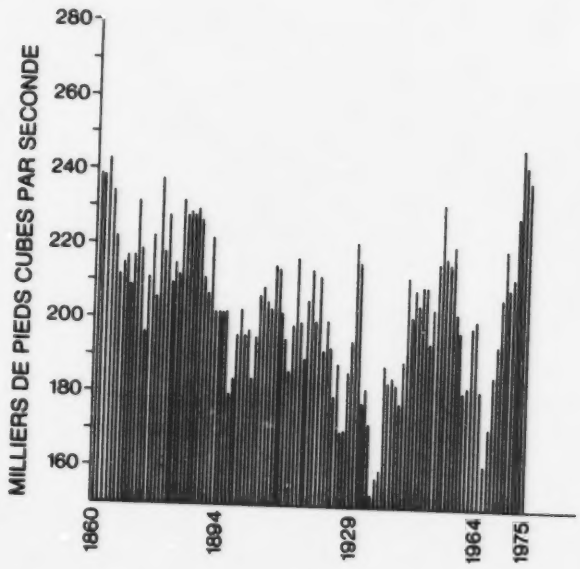



**Figure 20. Débit annuel moyen de la rivière Niagara à la hauteur de Queenston.**

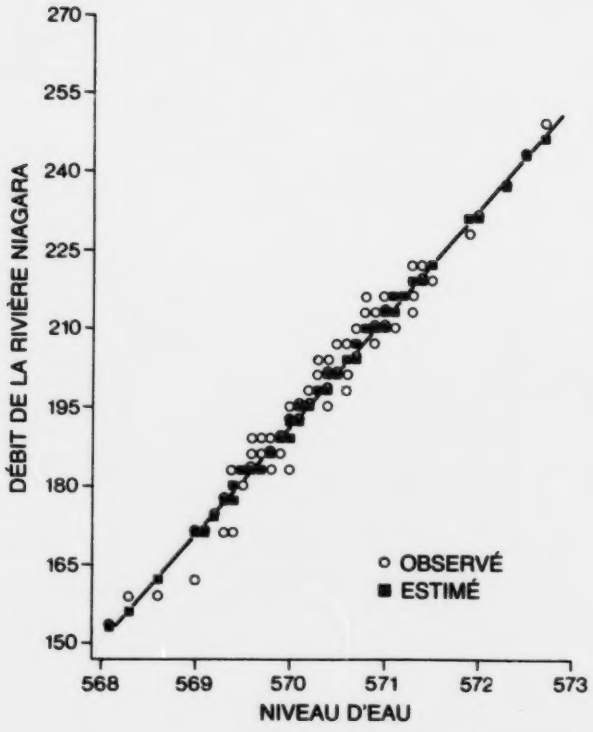



**Figure 21. Relation entre le débit de la rivière Niagara et le niveau du lac Érié.**

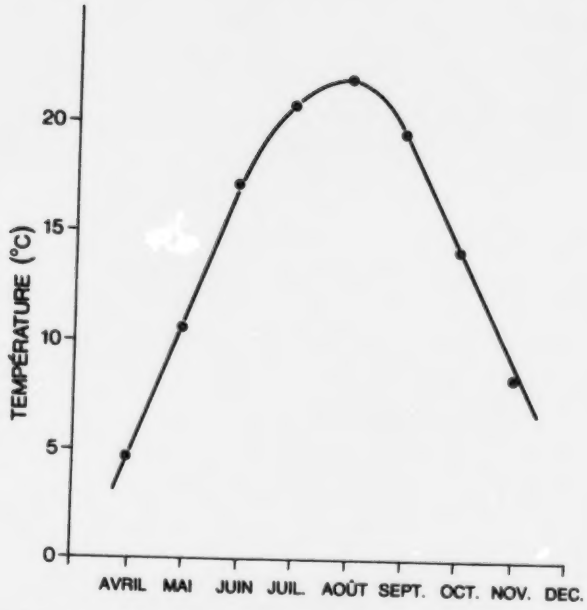



**Figure 22a.** Moyennes mensuelles des températures de l'air de 1967 à 1978 (bassin central).

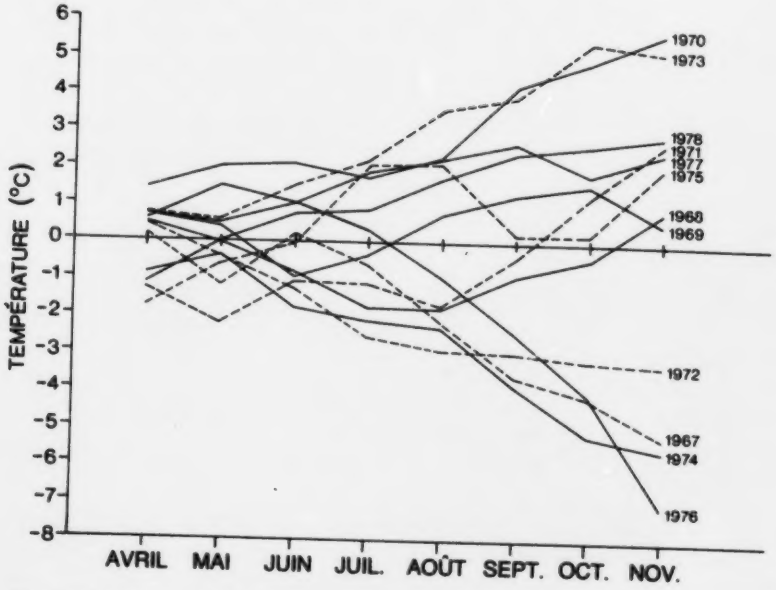



**Figure 22b.** Courbes des RC des températures de l'air de 1967 à 1978 (bassin central).

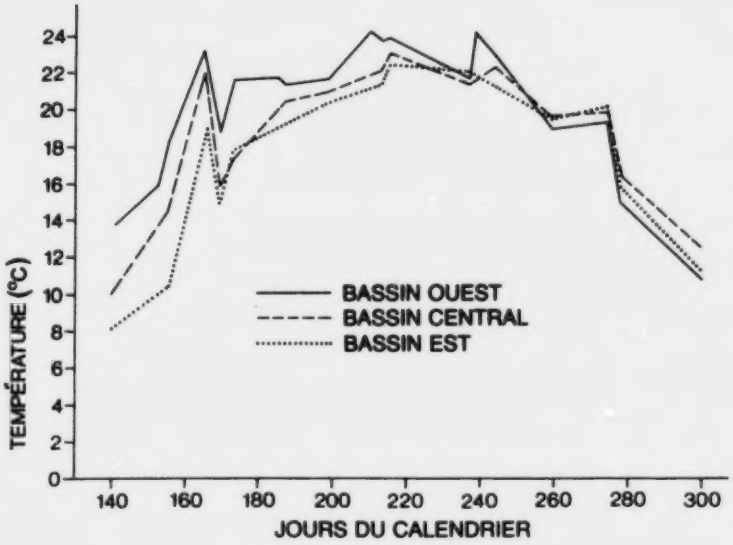



**Figure 23. Cycles saisonniers des températures dans les bassins ouest, central et est.**

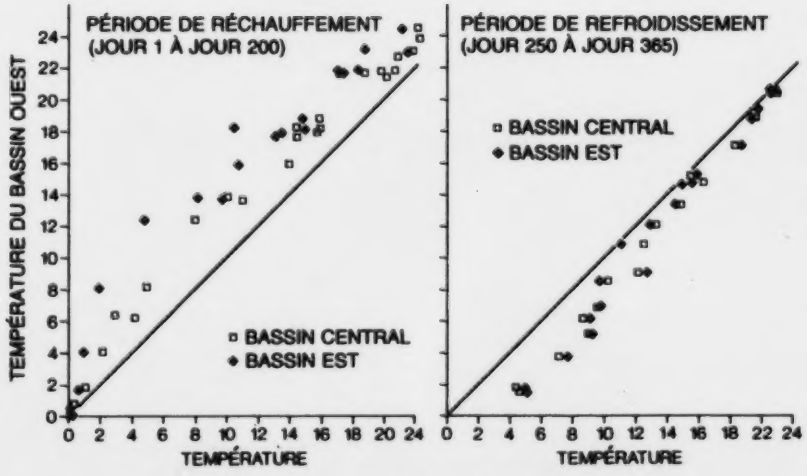



**Figure 24. Température moyenne de l'eau de surface du bassin ouest en fonction de celle des bassins central et est pendant les périodes de réchauffement et de refroidissement.**

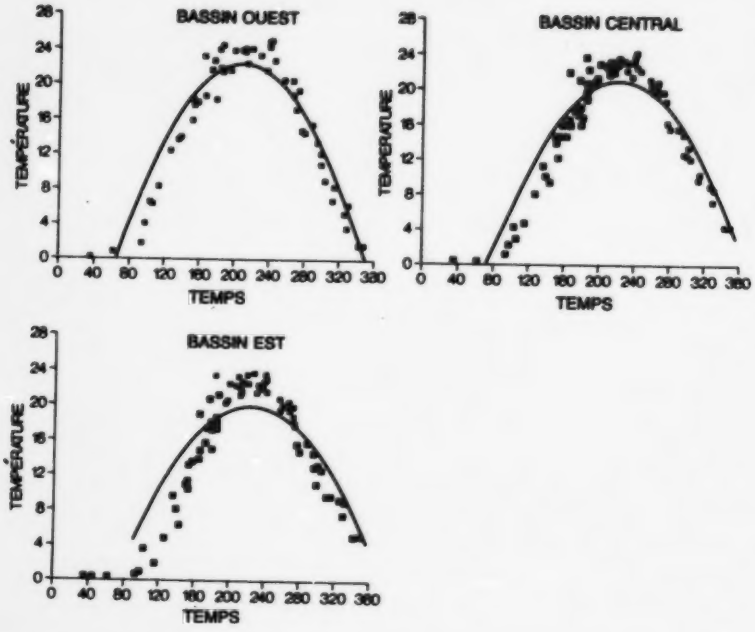



**Figure 25.** Tempérstures moyennes de l'eau de surface et valeurs estimées à l'aide du modèle 3.4.

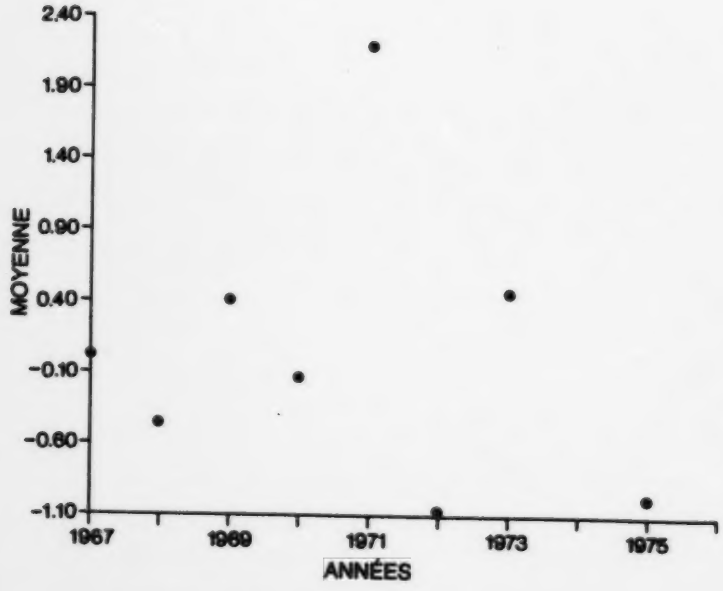



**Figure 26a.** Résidus moyens des tempérstures du bassin ouest.

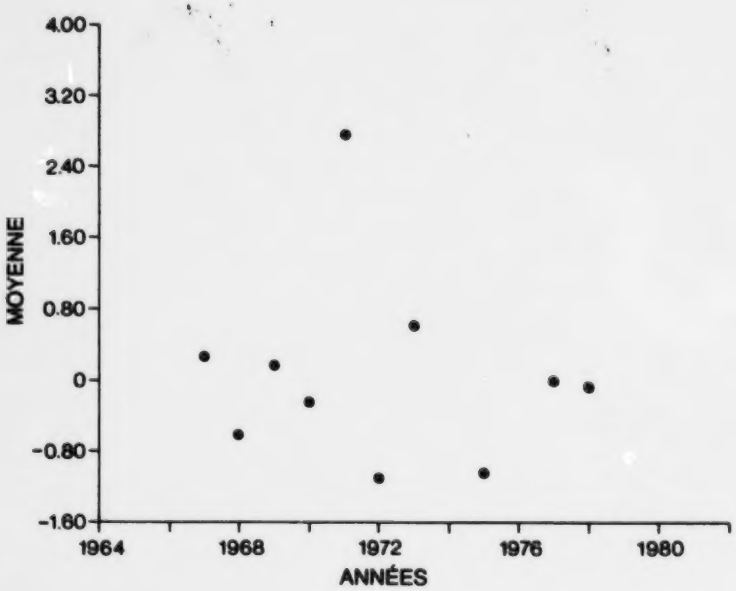



**Figure 26b. Résidus moyens des températures du bassin central.**

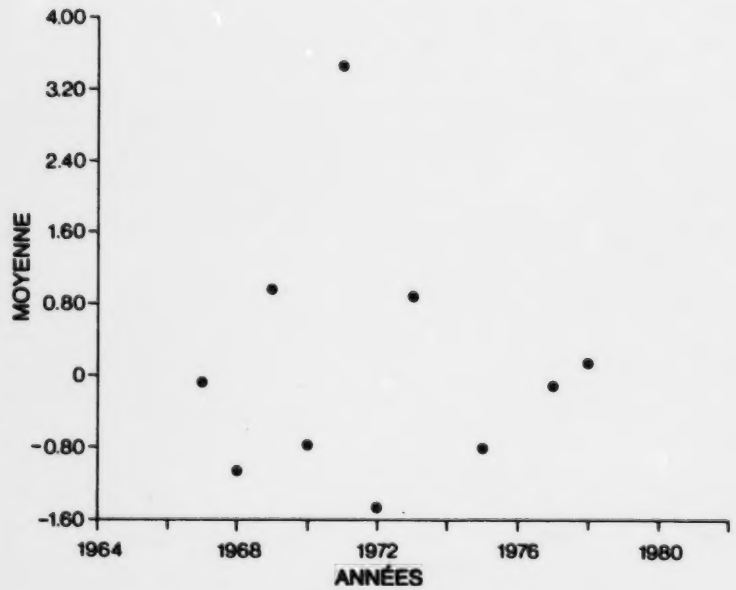



**Figure 26c. Résidus moyens des températures du bassin est.**

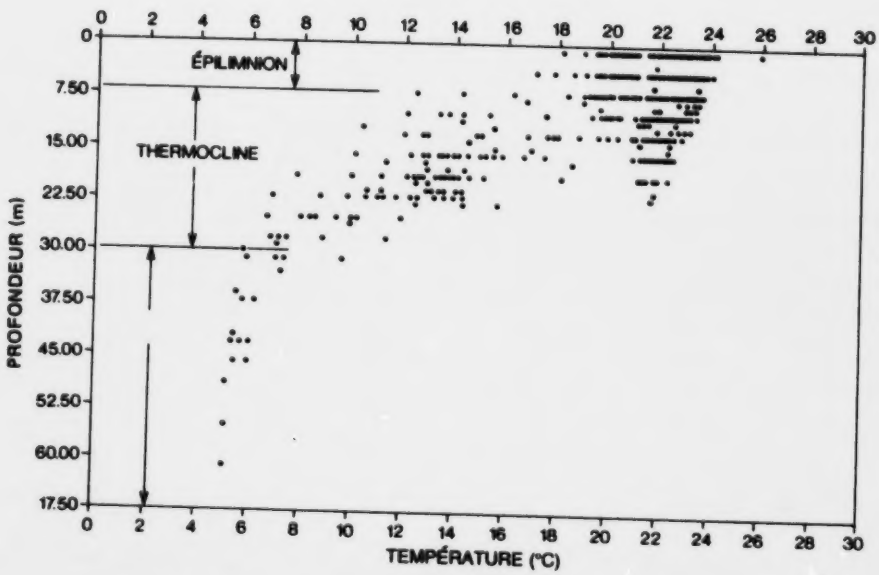



**Figure 27. Profil bathythermographique (29 juillet au 3 août 1968).**

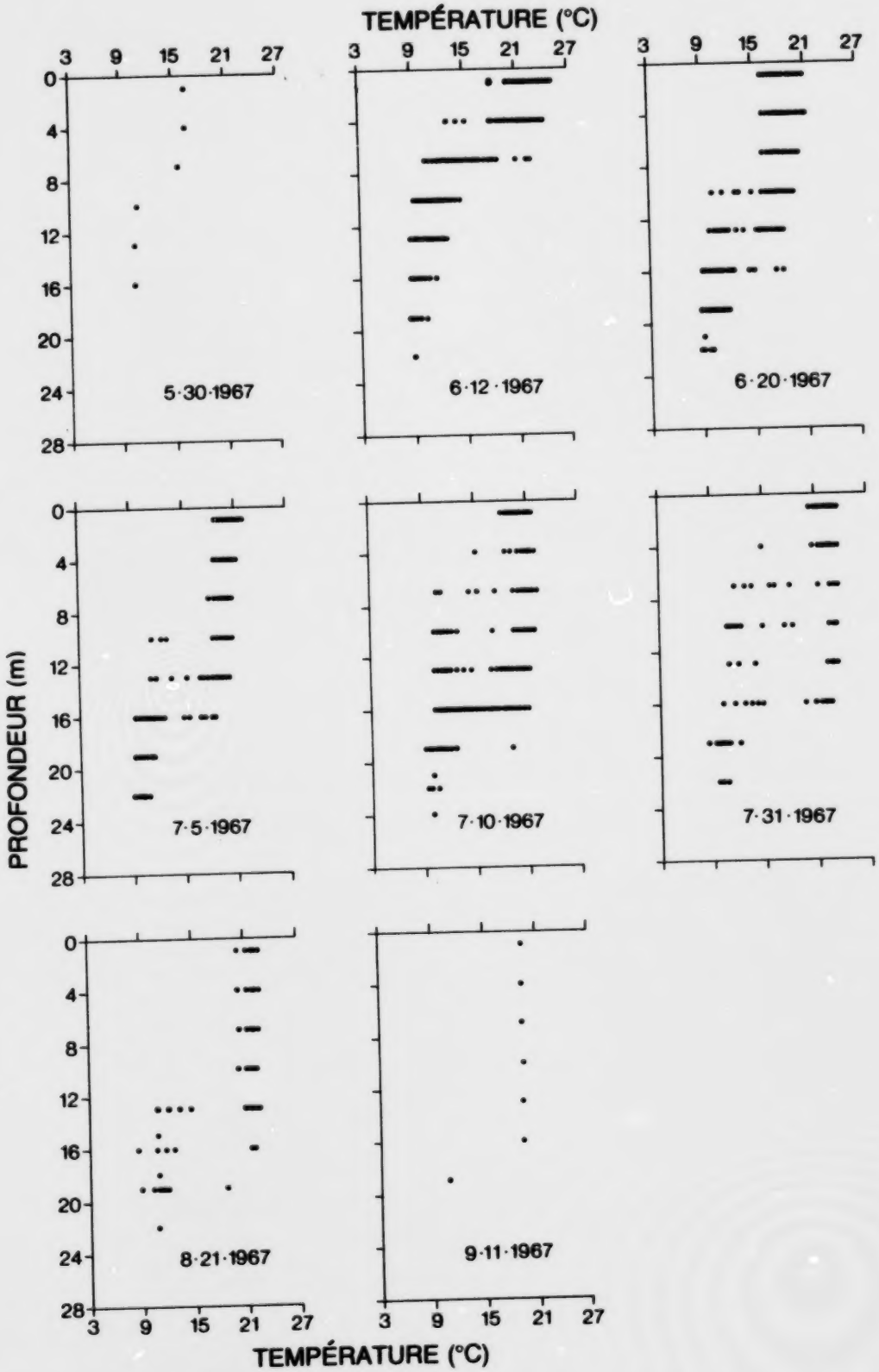



**Figure 28. Profil bathythermographique du bassin central, 1967.**

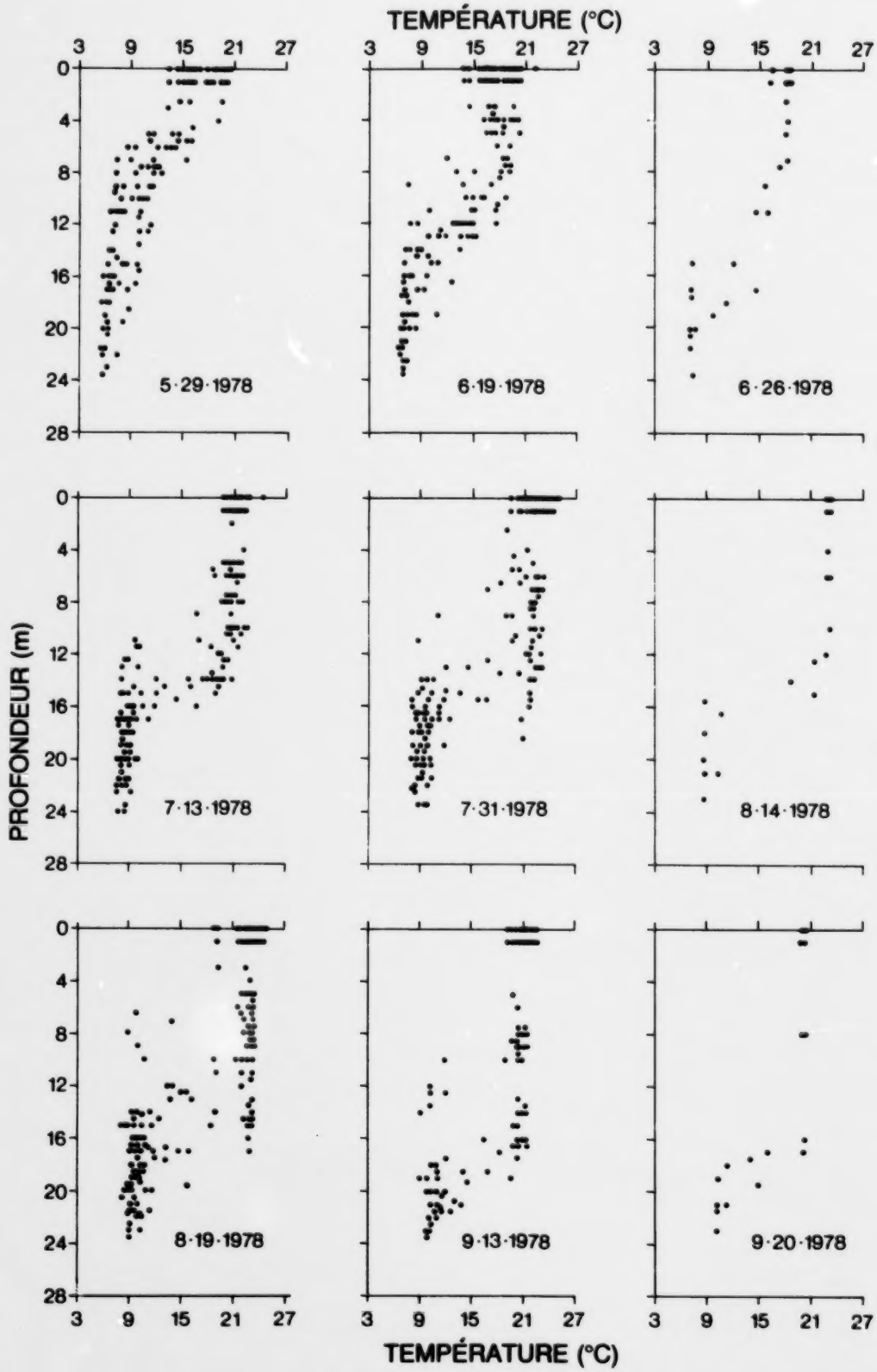



**Figure 29. Profil bathythermographique du bassin central, 1978.**

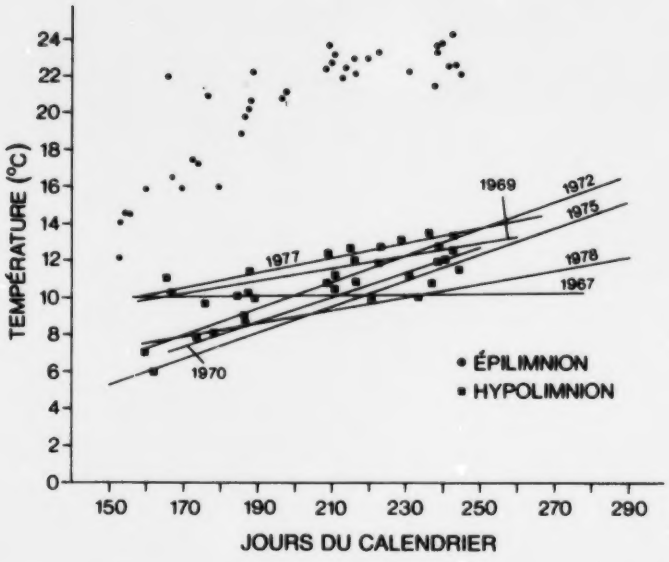



**Figure 30.** Températures moyennes de l'épilimnion et de l'hypolimnion du bassin central de 1967 à 1978 et droites de régression ajustées aux données sur l'hypolimnion.

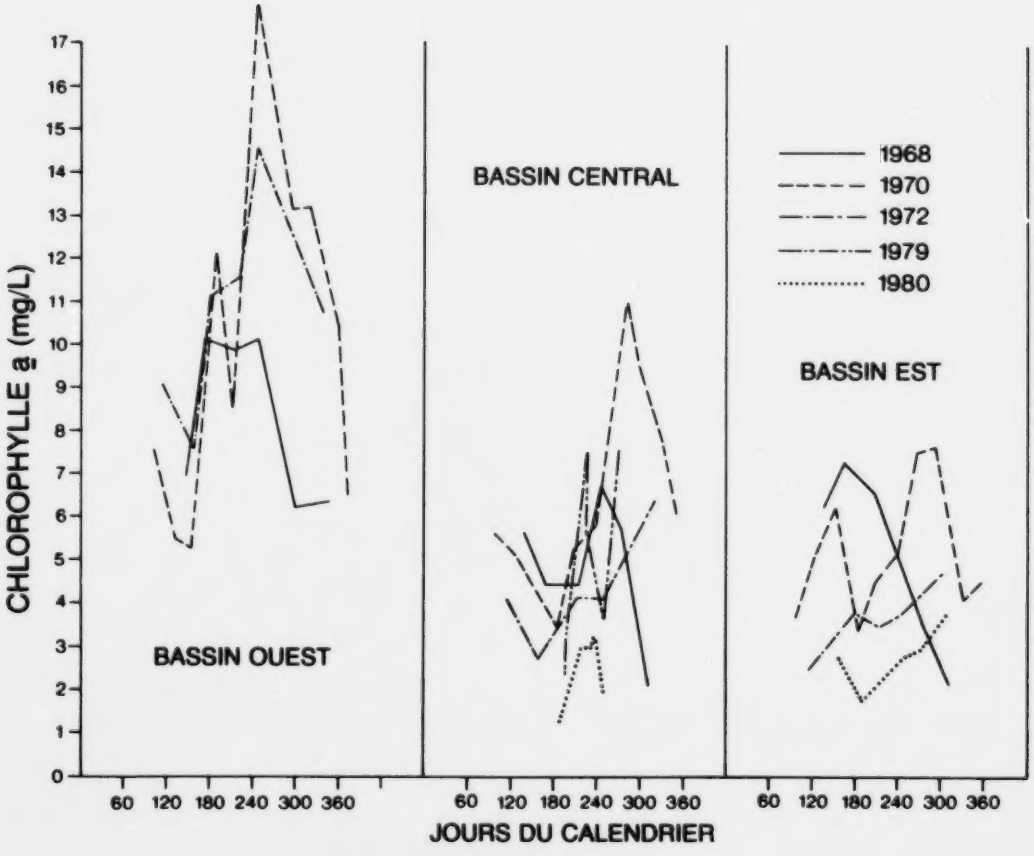



**Figure 31.** Teneurs moyennes non corrigées en chlorophylle *a* en fonction du temps.

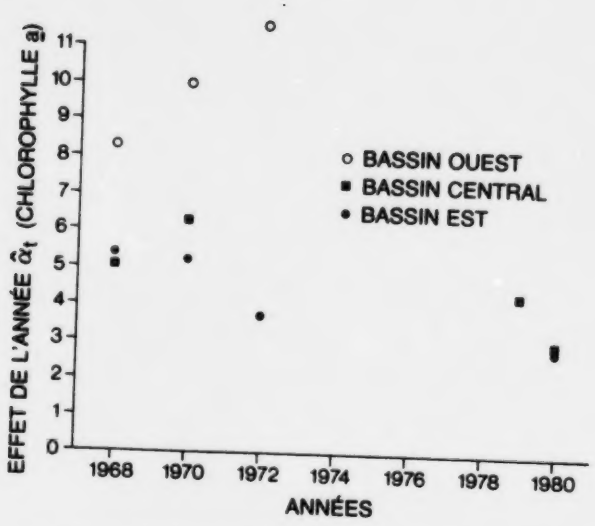



**Figure 32.** Estimations de l'effet de l'année en fonction des années (chlorophylle *a*).

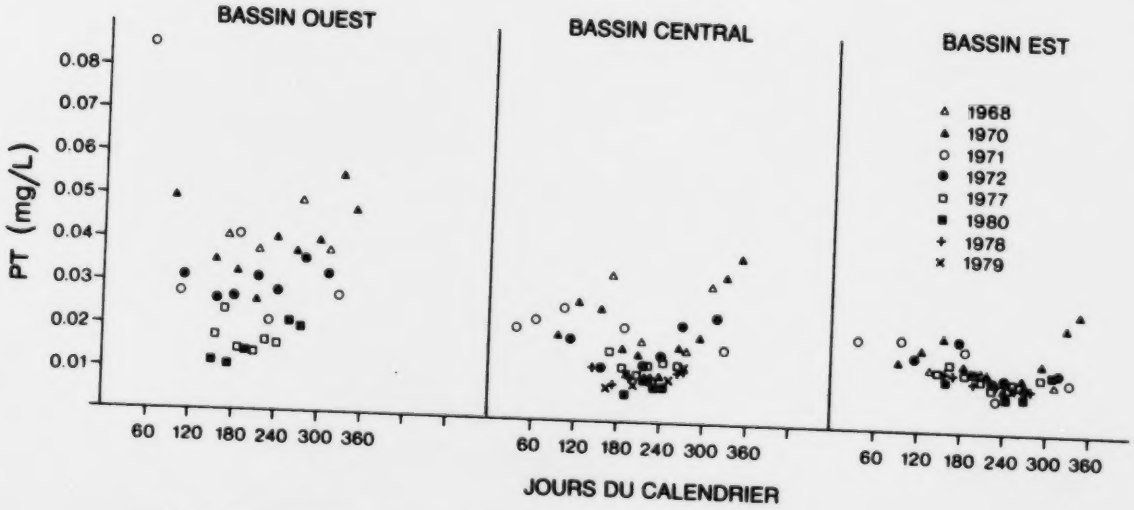



**Figure 33.** Teneurs moyennes en PT en fonction du jour de l'année dans les bassins ouest, central et est.

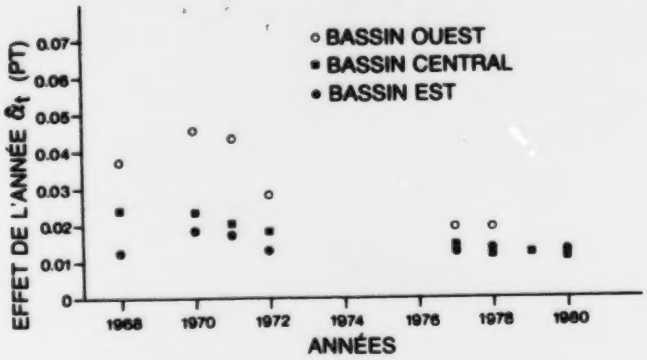



**Figure 34. Estimations de l'effet de l'année en fonction des années (PT).**

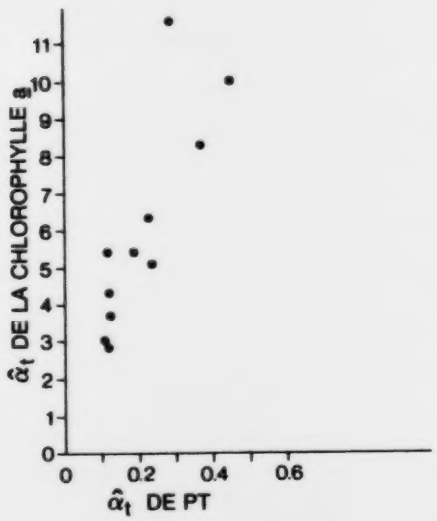



**Figure 35. Effet de l'année sur la chlorophylle *a* en fonction de celui sur le PT.**

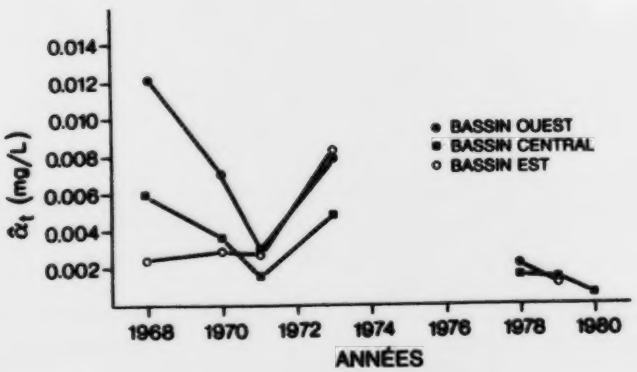



**Figure 36. Tendances des années sur le PRS.**

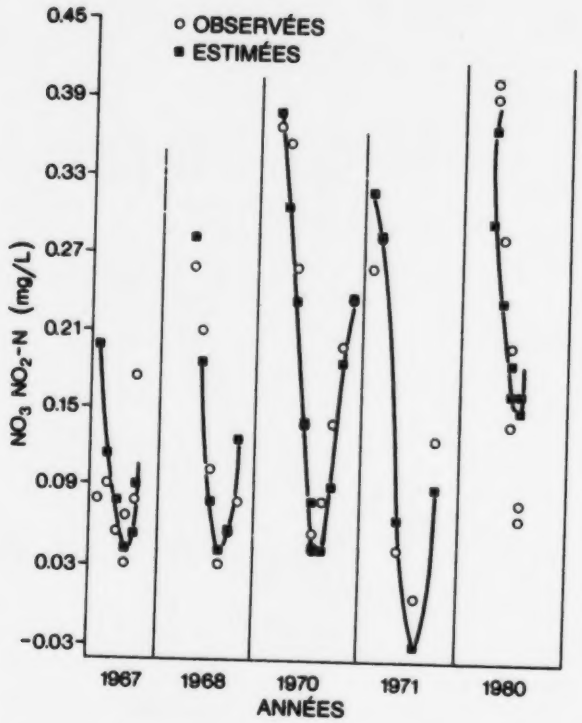



**Figure 37.** Teneurs en N-$NO_3$ $NO_2$ observées et estimées dans le bassin ouest.

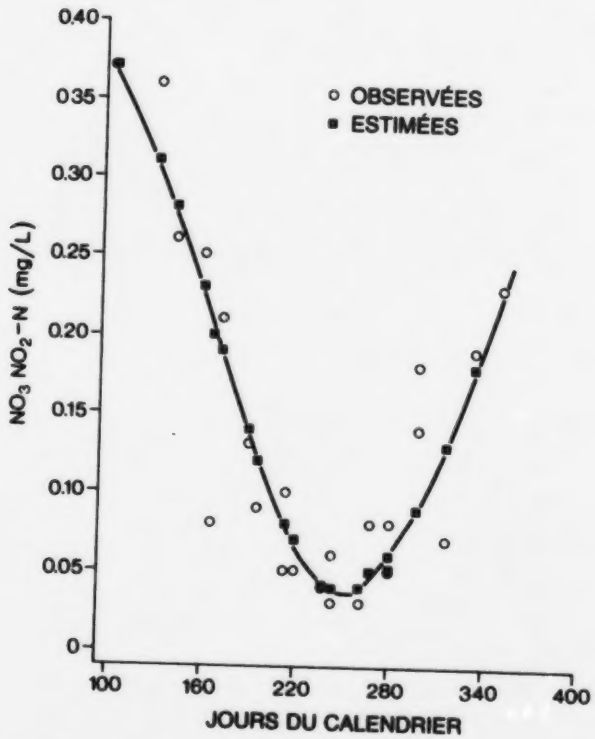



**Figure 38a.** Teneurs en N-$NO_3$ $NO_2$ observées et estimées dans le bassin ouest en 1967, 1968 et 1970.

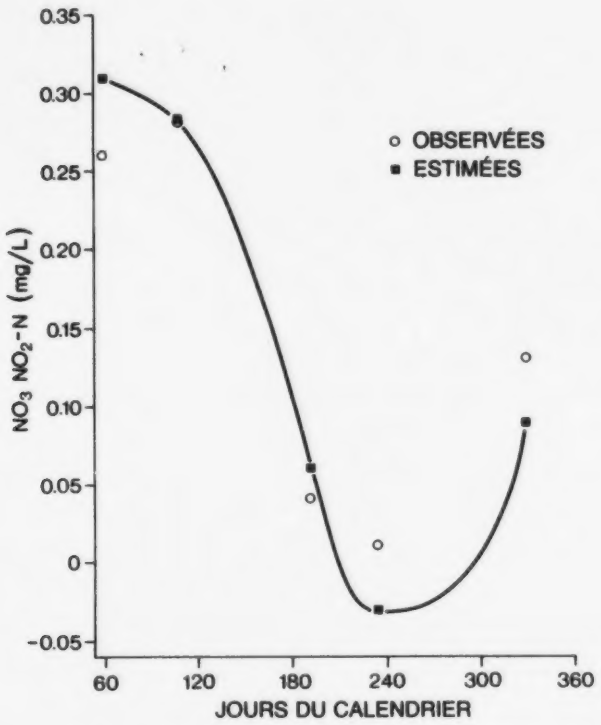



**Figure 38b. Teneurs en $N\text{-}NO_3NO_2$ observées et estimées dans le bassin ouest en 1971.**

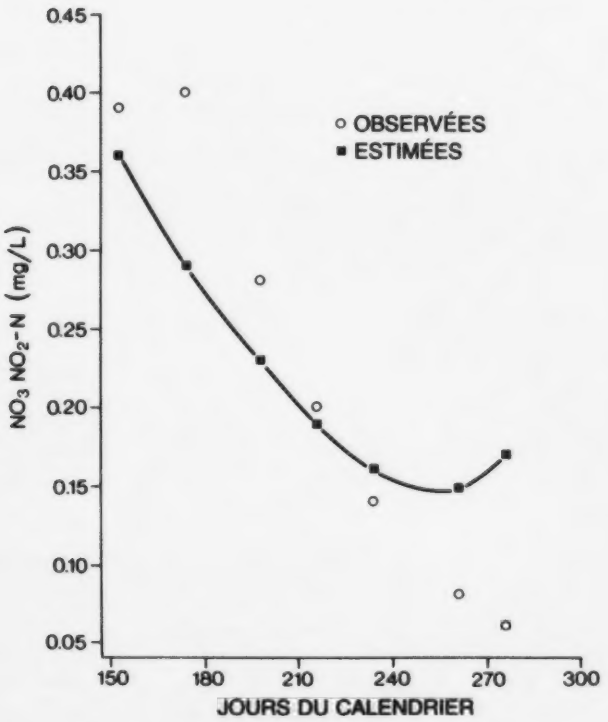



**Figure 38c. Teneurs en $N\text{-}NO_3NO_2$ observées et estimées dans le bassin ouest en 1978.**

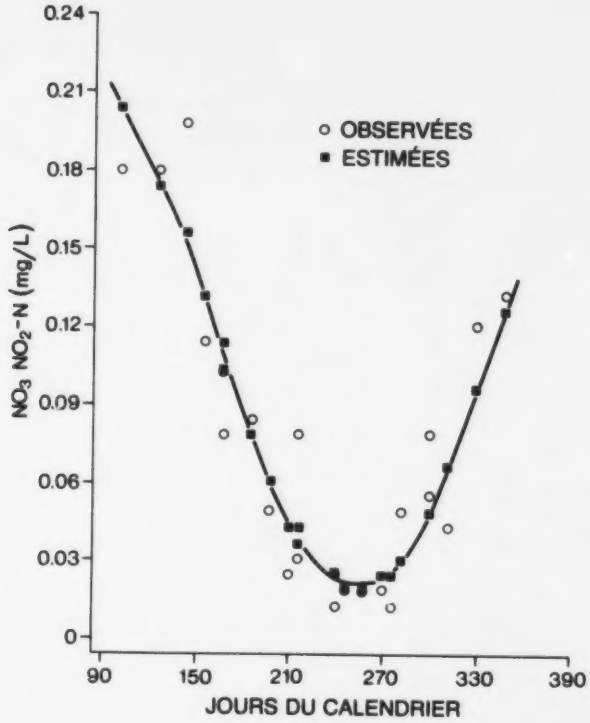



**Figure 39a.** Teneurs en N-$NO_3$,$NO_2$ observées et estimées dans le bassin central en 1967, 1968 et 1970.

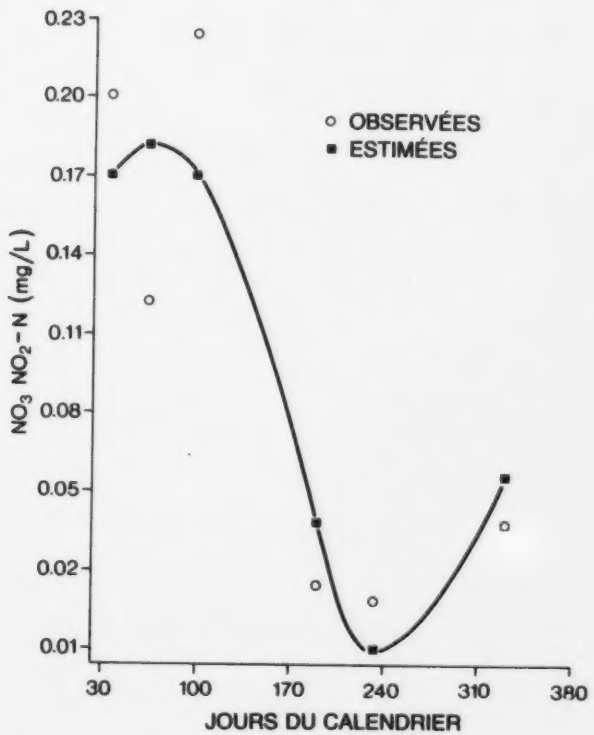



**Figure 39b.** Teneurs en N-$NO_3$,$NO_2$ observées et estimées dans le bassin central en 1971.

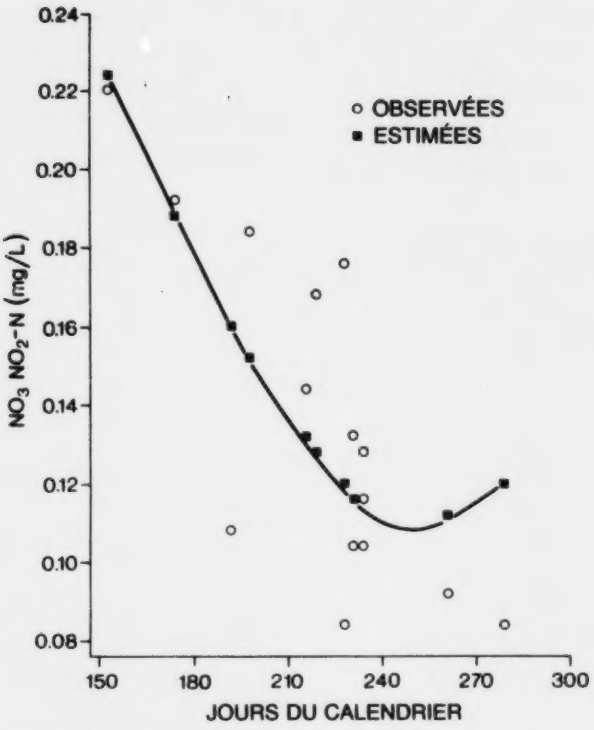



**Figure 39c.** Teneurs en N-$NO_3$ $NO_2$ observées et estimées dans le bassin central en 1978.

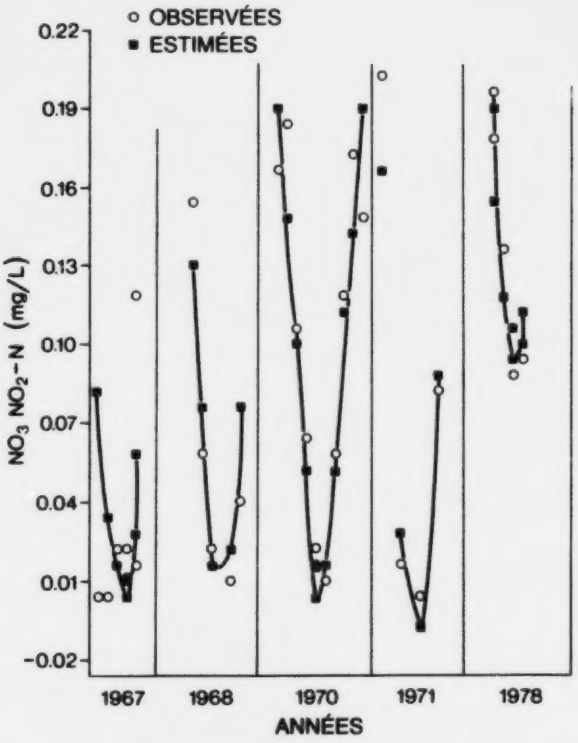



**Figure 40.** Teneurs en N-$NO_3$ $NO_2$ observées et estimées dans le bassin est.

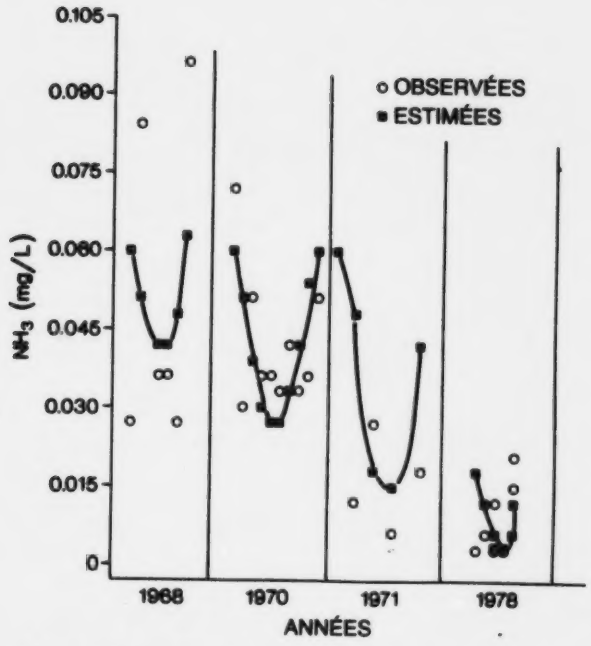



Figure 41. Teneurs en $NH_3$ observées et estimées dans le bassin ouest.

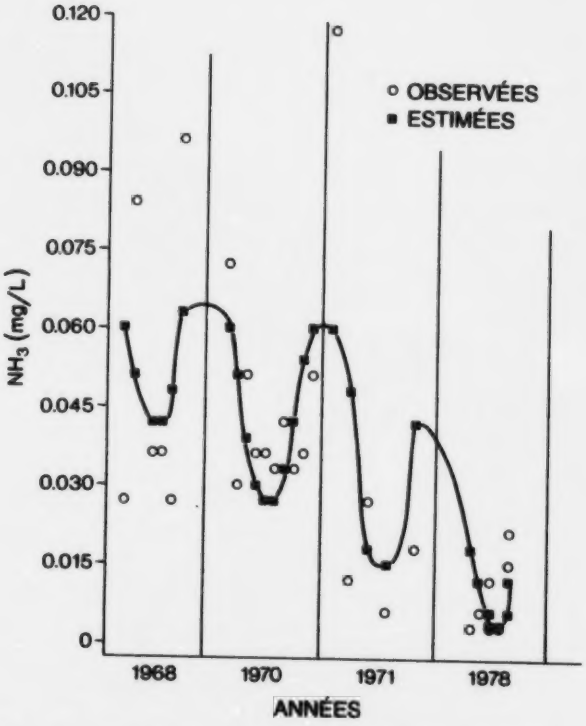



Figure 42. Teneurs en $NH_3$ observées et estimées dans le bassin central.

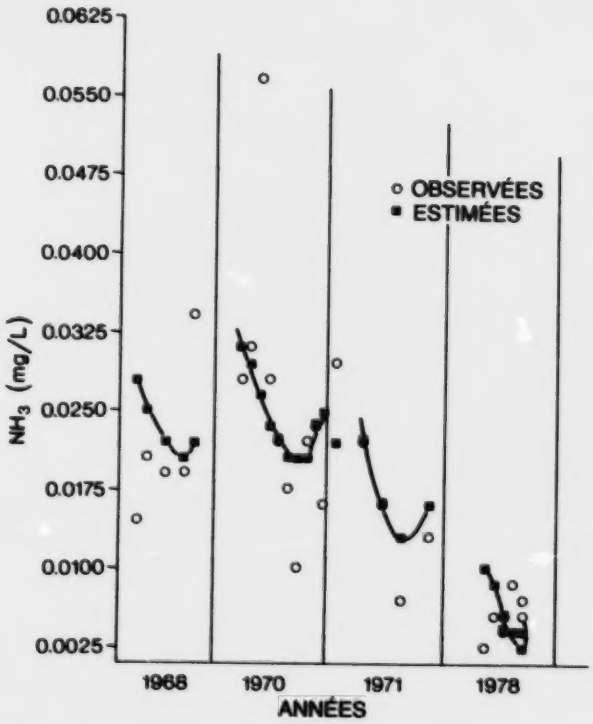



**Figure 43. Teneurs en $NH_3$ observées et estimées dans le bassin est.**

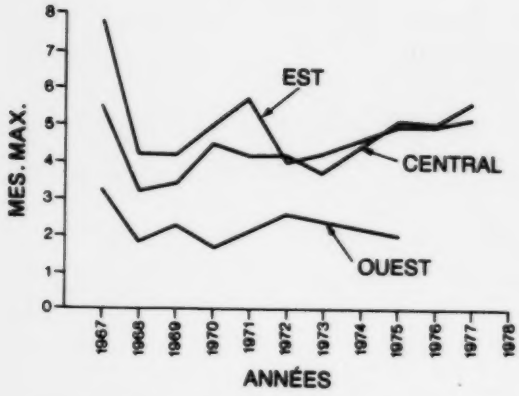



**Figure 44. Mesure de la profondeur estimée au disque de Secchi dans chacun des bassins en fonction de l'année.**

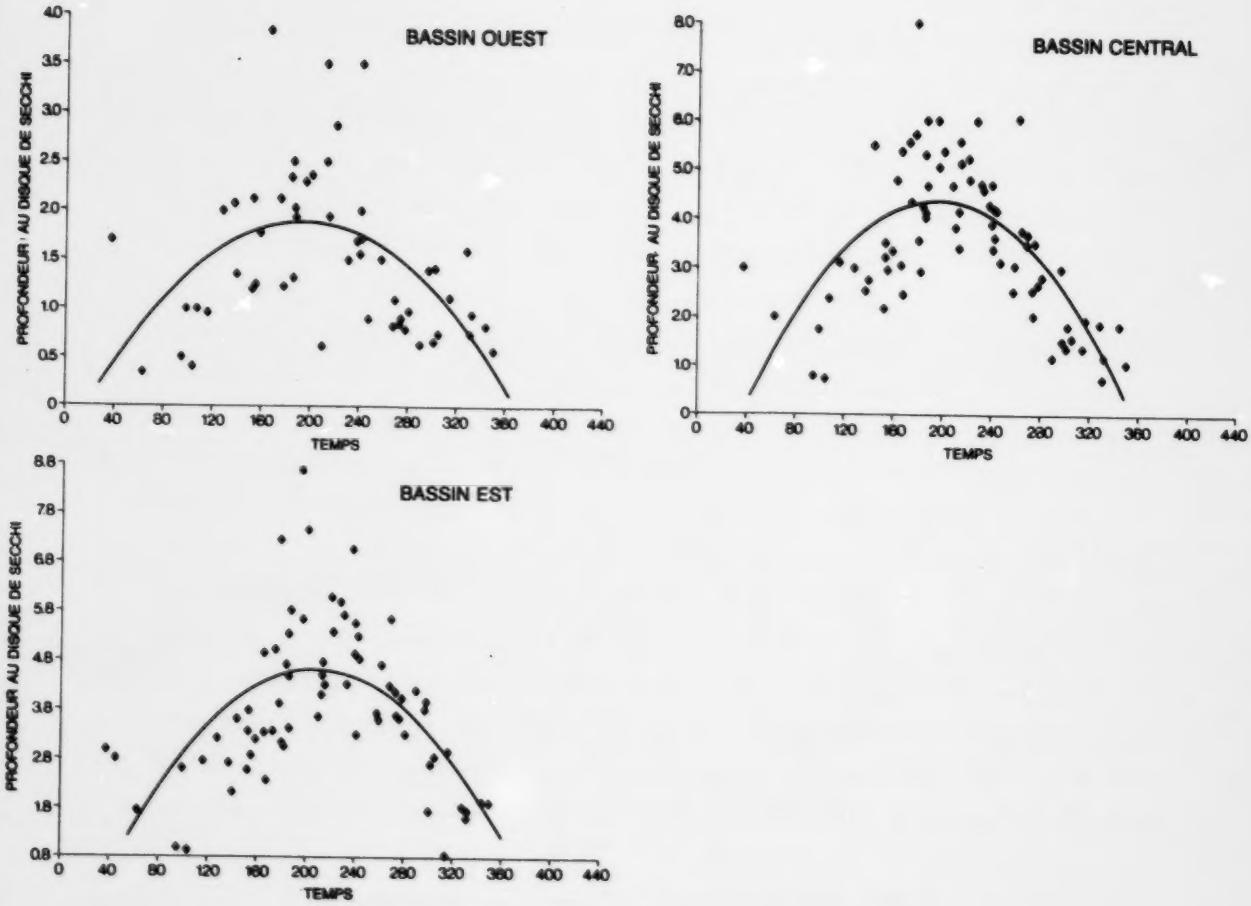



**Figure 45. Profondeurs observées au disque de Secchi et profondeurs estimées à l'aide du modèle 3.4.**

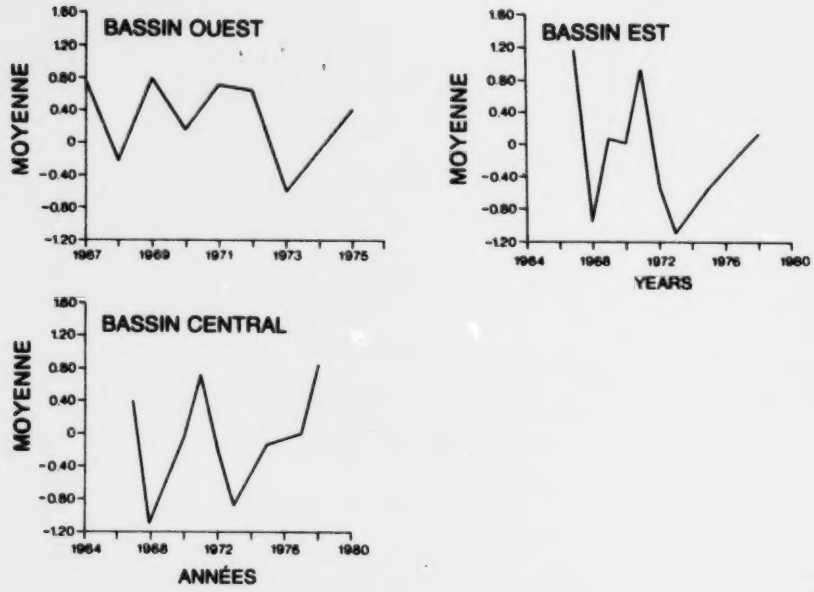



**Figure 46. Résidus moyens pour les profondeurs au disque de Secchi.**

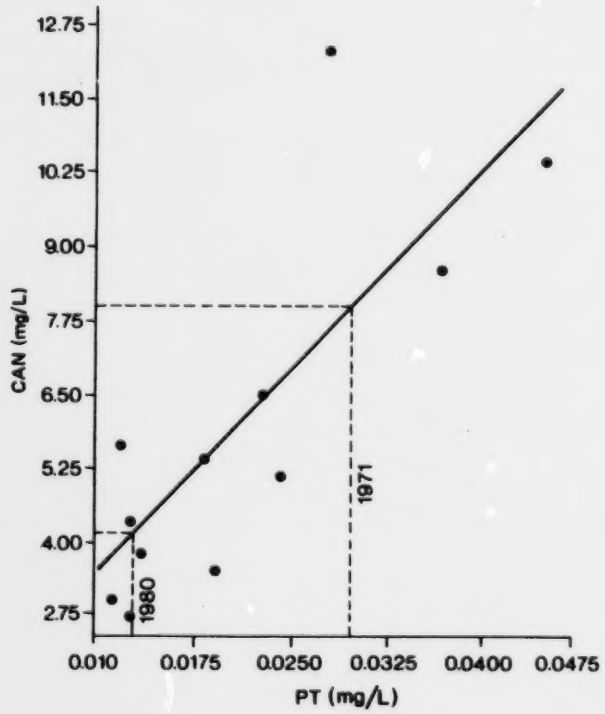



**Figure 47. Relation entre l'effet de l'année sur la chlorophylle *a* et l'effet de l'année sur le PT. CAN — Effets annuels non corrigés de la chlorophylle *a*.**

# Variabilité spatiale et temporelle de l'oxygène dissous dans le lac Érié

*par J.E. Anderson, A.H. El-Shaarawi, S.R. Esterby et T.E. Unny*


## RESUME

Une méthode non hiérarchique de regroupement en grappes autour des centroïdes les plus rapprochés a été utilisée pour répartir des paires de données (concentration d'oxygène dissous, température) en quatre groupes correspondant aux eaux hypolimnétiques et non-hypolimnétiques des bassins central et est. Pour les stations communes à toutes les missions d'une même année et classées comme faisant partie de l'hypolimnion, on a calculé la concentration initiale d'oxygène dissous et le taux d'appauvrissement en oxygène et on a testé la constance de ces paramètres au moyen d'une analyse de régression pondérée et de modèles de régression, la structure temporelle des données étant explicitement incorporée aux modèles. Les taux d'appauvrissement annuels non corrigés de 1967 à 1980 étaient semblables à ceux déjà signalés par plusieurs auteurs, ce qui montre que cette méthode semi-objective de regroupement en grappes constitue une solution de remplacement pratique au choix subjectif des données. On a montré que les conclusions d'une analyse de régression non pondérée sur la constance des concentrations et des taux d'appauvrissement différaient de celles d'une analyse de régression pondérée. Une analyse de régression avec poids arbitraires a révélé que ni la concentration initiale d'oxygène dissous ni le taux d'appauvrissement n'étaient demeurés constants

entre 1967 et 1980 dans le bassin central, mais que le taux d'appauvrissement était demeuré constant et que la concentration initiale d'oxygène dissous avait varié dans le bassin est.

## INTRODUCTION

La concentration d'oxygène dissous d'un lac est un indicateur de la qualité de l'eau et est utile à la classification des lacs, car elle joue un rôle de régulateur des processus métaboliques. L'appauvrissement en oxygène hypolimnétique des bassins central et est du lac Erié est considéré depuis longtemps comme un phénomène naturel qui débute au printemps et se termine à la fin de l'automne. Le gradient thermique apparait généralement au début de juin et marque le début de la période de stagnation estivale. La période de stratification du lac Erié dure environ 110 jours dans le bassin central et de 140 à 175 jours dans le bassin est. On ne note généralement pas de stratification persistante dans le bassin ouest. L'appauvrissement en oxygène de l'hypolimnion est caractéristique de la stratification d'été. Deux facteurs sont principalement responsables de ce phénomène. Tout d'abord, le remplacement de l'oxygène de l'hypolimnion est bloqué par le gradient de température et de densité de la thermocline. Deuxièmement, la décomposition de matières végétales et animales crée un dépôt additionnel de matières organiques dans la zone hypolimnétique. Comme des bactéries colonisent les sédiments et que leurs processus métaboliques de décomposition aérobie de la matière organique nécessitent de l'oxygène, la demande d'oxygène ne diminue pas. Ces conditions finissent par provoquer l'anoxie de certaines zones de l'hypolimnion. Ce processus est accéléré par l'apport de matières oxydables se présentant sous la forme de polluants ou d'autres substances produites par l'homme et se traduit, tous les autres facteurs demeurant constants, par une augmentation du taux d'appauvrissement en oxygène. Le taux d'appauvrissement en oxygène dissous est donc utile à l'évaluation des modifications de

l'environnement aquatique, surtout si l'on désire déterminer des tendances dans le temps.

Les taux d'appauvrissement en oxygène sont déterminés à partir des concentrations moyennes d'oxygène hypolimnétique en des temps différents. Le taux d'appauvrissement en oxygène dissous peut aussi être défini comme la demande d'oxygène nette par unité de volume (Fay et Herdendorf, 1981). Dans la présente étude, le taux d'appauvrissement en oxygène dissous est cependant un taux volumétrique calculé à partir des seules concentrations d'oxygène. Plusieurs méthodes ont, par le passé, été utilisées pour calculer le taux d'appauvrissement dans le lac Erié. Charlton (1979) et Dobson et Gilbertson (1971) ont calculé les taux d'appauvrissement en oxygène dissous dans les zones hypolimnétiques du lac Erié en faisant appel à une technique simple de régression linéaire; la concentration d'oxygène moyenne de l'hypolimnion de tout le bassin a été établie par régression en fonction du temps pour le point milieu de chaque mission. Ce mode d'analyse a toujours été le plus répandu. Le modèle d'échange dans le mésolimnion, élaboré par Burns (1976a), traite de façon quantitative des eaux hypolimnétiques des deux bassins et de leur teneur en oxygène dissous. Il tient aussi compte des apports et des pertes d'oxygène dans un bassin par circulation verticale ou horizontale et comporte des corrections pour les diverses températures de l'hypolimnion.

Burns et Rosa (1981) ont tenté de délimiter une zone homogène représentative de chacun des bassins en établissant une carte de distribution des taux d'appauvrissement; leur analyse fait cependant appel à des taux par intervalle. Un taux d'appauvrissement par intervalle est le rapport de la perte d'oxygène (notée à une station ou exprimée sous forme d'écart par rapport à la moyenne dans une zone représentative) au temps (en jours) séparant deux missions. Dans le cas d'une période de stratification comprenant NC missions (NC-1), on pondère les taux d'appauvrissement par intervalle de façon à obtenir un taux général pour toute la période de stratification.

Il n'y a pas que les méthodes de calcul des taux d'appauvrissement qui soient différentes, les critères utilisés pour déterminer la concentration d'oxygène hypolimnétique dissous à une station d'échantillonnage varient aussi. Dobson et Gilbertson (1971) se sont servis de la moyenne arithmétique des concentrations des échantillons pour lesquels la température s'écartait de moins de 3°C de la température minimale. Charlton (1979) n'a utilisé que les valeurs obtenues près du fond aux stations où la profondeur était supérieure à 15 m et où il y avait stratification et absence d'indice d'une arrivée d'eau du bassin est; il a éliminé les données qu'il considérait erronées pour les raisons présentées à l'appendice de sa publication. Ces deux publications ne traitent que des taux d'appauvrissement dans le bassin central. Burns et Rosa (1981) ont choisi des valeurs en fonction de la température et de la profondeur d'échantillonnage. Ces auteurs, en plus de Carr (1962) et Beeton (1963) conviennent que les problèmes d'échantillonnage et les prélèvements non représentatifs imposent un certain choix dans les données. Il ressort finalement que les conclusions relatives au taux d'appauvrissement en oxygène dans l'hypolimnion, tirées depuis 1929, ne concordent pas.

Le présent chapitre traite du problème du choix des données et de méthodes permettant d'estimer les taux d'appauvrissement et les concentrations initiales ainsi que de tester l'hypothèse de la constance dans le temps. On propose une méthode de regroupement en grappes comme approche semi-objective et pratique à la détermination d'ensembles homogènes d'observations relatives aux eaux hypolimnétiques. Les concentrations hypolimnétiques moyennes d'oxygène dissous sont calculées pour les groupes homogènes, ce qui réduit la variabilité due à la variation spatiale des concentrations et aux échantillonnages non représentatifs. Une régression linéaire qui tient compte de la précision des moyennes des groupes sert à estimer les taux d'appauvrissement et à tester la constance dans le temps; on utilise, pour ce faire, des modèles tenant compte de la structure temporelle des données.

## DONNEES ET METHODES D'ANALYSE

On a utilisé les données recueillies de 1967 à 1980 par le Centre canadien des eaux intérieures. La concentration d'oxygène dissous a été déterminée par la méthode de Winkler (p. ex. Wetzel et Likens, 1979). On a choisi d'utiliser la température de l'eau de pair avec la concentration d'oxygène dissous, l'hypolimnion étant défini en fonction de la température, et seules les valeurs appariées surface-fond ont été retenues pour chaque station. De plus, on ne s'est intéressé qu'aux seules missions effectuées pendant la période de stratification d'été. Il était donc prévu que les groupes caractérisant l'hypolimnion des bassins central et est seraient séparés de ceux représentant les eaux étrangères à l'hypolimnion et que les données sur les eaux du fond aux stations sans stratification seraient incorporées à celles sur les groupes étrangers à l'hypolimnion.

### Méthode de regroupement en grappes

Une méthode non hiérarchique de regroupement en un nombre variable de grappes autour des centroïdes les plus rapprochés a été utilisée. L'algorithme de regroupement en grappes (Anderson, 1982) s'inspire de l'algorithme ISODATA bien connu de Ball et Hall (1967). On y retrouve en effet bon nombre d'étapes identiques, mais les méthodes utilisées au sein de chaque étape diffèrent considérablement. Les étapes du présent algorithme, résumées à la figure 48, comprennent la génération automatique de germes ainsi que des critères d'élimination des grappes trop clairsemées, de scindage des grappes allongées et de regroupements des grappes trop rapprochées.

Soit $X = \{x_{jj}\}$ la matrice des données où $x_{ij}$ est la valeur observée de la $j^{ième}$ variable dans la $i^{ième}$ unité; $i = 1,2...n$ et $j = 1,2...m$. Les unités, représentées par les

vecteurs $x_i = (x_{i1}, x_{i2} \ldots x_{im})'$ pour i = 1,2...n seront donc assignées aux grappes appropriées. Les étapes de l'algorithme sont décrites plus bas.

1. Initialisation et transformation. L'utilisateur précise le nombre initial de grappes (k) ainsi que le nombre minimal de membres par grappe. La première étape de calcul est une transformation facultative des observations qui consiste à diviser chaque observation par l'écart type de la variable, afin d'éliminer l'effet d'unités de mesure différentes. Ainsi, $x_i$ devient

$$x'_{ij} = x_{ij}/s_j$$

où $$s_j^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_{ij} - \bar{x}_{.j})^2 \text{ et } \bar{x}_{.j} = \frac{1}{n} \sum_{i=1}^{n} x_{ij}$$

Les observations, transformées ou non transformées, seront, dorénavant, représentées par $x_{ij}$. Les centroïdes initiaux des grappes, ou germes $z_\ell$ où $\ell = 1,2 \ldots k$ sont donnés par le calcul suivant (Northouse et Fromm, 1976):

$$z_{lj} = \bar{x}_{.j} + \text{SGN} \{\sin (\frac{2\pi\ell}{2^j} - \frac{2\pi}{2^{j+1}})\} s_j$$

où $\ell = 1,2 \ldots k$ et $j = 1,2 \ldots m$ et où $\bar{x}_{.j}$ et $s_j$ sont la moyenne et l'écart type de la $j^{\text{ième}}$ variable, transformée ou non transformée.

2. Affectation et recalcul des centroïdes. Chaque unité, $x_i = (x_{i1}, x_{i2} \ldots x_{im})'$, est affectée à la grappe pour laquelle la distance euclidienne entre l'unité et le centroïde de la grappe est minimale. Une fois toutes les unités affectées à l'une des k grappes, les centroïdes des grappes sont réévalués. Soit

$\omega_\ell$ l'ensemble des unités formant la $\ell^{ième}$ grappe; on obtient alors

$$z_{\ell j} = \frac{1}{n_\ell} x_{ij} \sum_{\epsilon\omega_\ell} x_{ij}$$

où $n_\ell$ est égal au nombre d'unités dans la grappe $\ell$.

3. Critère de l'insuffisance de membres. Toute grappe contenant moins de membres que le nombre minimum précisé par l'utilisateur est éliminée. Les membres éliminés sont réaffectés, et les centroïdes des grappes, recalculés de la façon décrite à l'étape 2 ci-dessus.

4. Critère de fractionnement. Le rapport d'une mesure de la dispersion d'une grappe non fractionnée à la somme des mesures de dispersion de deux sous-grappes est comparé à un rapport seuil ($\theta_s$) précisé par l'utilisateur, ceci pour déterminer si la grappe devrait être fractionnée (Duda et Hart, 1973). La matrice de dispersion pour la grappe $\ell$ est donnée par

$$S_T = \sum_{\underline{x}_i \epsilon\omega_\ell} (\underline{x}_i - z_\ell)(\underline{x}_i - z_\ell)'$$

où $z_\ell = (z_{\ell 1}, z_{\ell 2} \cdots z_{\ell k})'$

On détermine, à l'aide de la matrice de dispersion, l'axe principal et on pratique le long de cet axe trois coupes également espacées pour diviser la grappe en paires de sous-grappes. La mesure de dispersion au sein d'une grappe est donnée par la somme des sommes des carrés intra-grappes pour toutes les variables; on l'appelle la trace de la matrice de dispersion et on la désigne par $tr(S_T)$ pour la grappe d'origine. Dans le cas des paires de sous-grappes, on

additionne les sommes des carrés intra-grappes des deux grappes que l'on exprime par $tr(S_w)$. La paire de sous-grappes présentant la $tr(S_w)$ la plus petite sert à tester si la réduction a été suffisante au sein de la dispersion intra-grappe. Si la grappe est

$$\frac{tr(S_T)}{\min tr\ (S_w)} > \theta_s$$

fractionnée en la paire de sous-grappes correspondant à la $tr(S_w)$ minimale et les centroïdes des nouvelles sous-grappes sont déterminés. Ce critère de fractionnement est appliqué indépendamment à chacune des grappes de départ.

5. Critère de fusionnement. Le paramètre de fusionnement $\rho_K$, défini par Northouse et Fromm (1976), est

$$\rho_K = \frac{m}{3K} \sum_{k=1}^{K} D_k$$

où m est le nombre de variables, K le nombre actuel de grappes et $D_k$ la distance minimale entre la grappe k et les k-1 autres grappes. Si $D_k < \rho_K$, la $k^{ième}$ grappe est éliminée. Les membres de la grappe éliminée sont ensuite réaffectés et les centroïdes sont recalculés selon l'étape 2.

6. Test de stabilité. La configuration définitive est atteinte quand le nombre de grappes est identique au nombre obtenu à l'itération précédente ou quand le nombre maximal d'itérations a été dépassé. Dans les autres cas, les étapes 2 à 5 sont répétées jusqu'à l'obtention d'une configuration définitive. Une fois les grappes définitives choisies, la moyenne et l'écart-type des concentrations d'oxygène dissous sont calculés pour chacune des grappes.

Application de la méthode de regroupement en grappes

Dans le présent contexte, $(x_i = x_{11}, x_{12})'$ où $x_{11}$ est la température et $x_{12}$ la concentration d'oxygène dissous à une station et une profondeur données. Ainsi, $x_i$ pourrait être la paire de données obtenue en surface et $x_j$ celle obtenue au fond à une certaine station. Pour une mission particulière, toutes les données disponibles sur les bassins central et est formeraient la matrice X et seraient donc réparties en grappes.

La méthode de regroupement en grappes est décrite dans un programme FORTRAN; on pourra trouver à la figure 49 un exemple de l'information graphique obtenue après un passage type du programme, celui de la mission 7022106. La présentation graphique des données sous forme de numéros identifiant les grappes auxquelles les points ont réellement été assignés permet d'évaluer visuellement l'efficacité et l'évolution du regroupement en grappes. Les grappes 1 et 4 correspondent aux données présentant les caractéristiques de l'épilimnion, à savoir les eaux les plus chaudes et les concentrations d'oxygène dissous les plus élevées. La grappe 2 regroupe les concentrations d'oxygène dissous relativement élevées de l'hypolimnion du bassin est, où la température de l'eau est la plus froide. Quant à la grappe 3, elle regroupe les teneurs en oxygène de l'hypolimnion du bassin central. On pourra noter sur une carte des emplacements des stations formant les grappes 2 et 3, les grappes de données que l'on considère comme faisant partie de l'hypolimnion. En plus de ces graphiques, le programme imprime des statistiques sur les grappes définitives, à savoir: la moyenne et l'écart-type des concentrations d'oxygène dissous, la température et la profondeur, ceci pour chacune des grappes.

Le programme informatique est interactif et certaines corrections ont été apportées afin d'atteindre les objectifs qui étaient: (1) la séparation des grappes regroupant des stations des

bassins est et central, (2) le fusionnement de deux grappes de l'hypolimnion du bassin central, et (3) la réaffectation de points isolés mal classifiés (Anderson, 1982). La méthode de regroupement en grappes décrite plus haut a été appliquée aux données des années 1967 à 1970. Les résultats obtenus, ainsi que les mesures de profondeur, ont servi à la classification pour les années 1971 à 1980. De plus, on n'a utilisé que les stations communes à toutes les missions pour une même année. Cela a eu pour effet de limiter le nombre de points dans la plupart des grappes et d'exclure les années 1971, 1974, 1976 et 1980 (bassin central), le nombre de points retenus étant insuffisants. Les grappes définitives ont servi à calculer les moyennes et les écarts-types des concentrations d'oxygène dissous devant servir à déterminer les taux d'appauvrissement.

## Taux d'appauvrissement - Calculs et tests

### Calcul des taux d'appauvrissement

Les observations montrent que la concentration moyenne d'oxygène dissous de l'hypolimnion décroît de façon linéaire tout au long de la stratification. Il convient donc d'utiliser le modèle

$$y_i = \alpha_0 + \alpha_1 x_i + \varepsilon_i \quad i = 1,2...NM \qquad (4.1)$$

où
- $NM$ = le nombre de missions effectuées dans une zone donnée au cours de l'année étudiée
- $y_i$ = la concentration moyenne d'oxygène dissous (mg/L)
- $x_i$ = la date correspondante à mi-mission (jours du calendrier $1^{er}$ avril)
- $\varepsilon_i$ = la variable aléatoire de moyenne 0
- $\alpha_0$, $\alpha_1$ = les paramètres inconnus à estimer à partir des données.

Le paramètre $\alpha_1$ représente le taux d'appauvrissement. La méthode généralement utilisée pour estimer $\alpha_0$ et $\alpha_1$ est celle des moindres carrés non pondérés. Cela équivaut à supposer que la distribution de $\varepsilon_1$ est normale, de moyenne 0 et de variance $\sigma^2$. Cette méthode convient si le nombre de stations servant à l'analyse demeure constant, non seulement au cours de l'année, mais tout au long de la période d'étude et que l'écart-type de la concentration d'oxygène dissous est indépendant du jour (c'est-à-dire du moment de la cueillette des données). Le nombre de stations d'échantillonnage, constant au sein des années, varie cependant d'une année à l'autre. De plus, l'écart-type varie selon le jour du calendrier, ceci chaque année, et selon le nombre de stations d'échantillonnage de la zone hypolimnétique. Pour le démontrer, on a établi une droite de régression simple de l'écart-type (variable dépendante) de la concentration d'oxygène dissous en fonction du nombre correspondant de stations d'échantillonnage pour les données relatives à l'hypolimnion (variable indépendante) du bassin central. Les résultats ont montré l'existence d'une corrélation positive fortement significative ($[P < 0.01]$ et $F_{1,40} = 11.39$) entre l'écart-type, y, et le nombre de stations, x. L'équation de régression est

$$y = 0.654 + 0.030x$$

L'hypolimnion n'est donc pas totalement homogène et le nombre de stations d'échantillonnage disponibles est insuffisant pour permettre de tirer une conclusion par analyse exacte. La régression a été répétée en fonction du nombre de stations et du jour du calendrier. L'équation de régression est

$$y = 1.206 + 0.33x - 0.086z$$

où z est le jour du calendrier. L'inclusion de z s'est avérée fortement significative ($P \leq 0.01$ et $F_{1,39} = 12.23$). On pourra trouver à la figure 50 le report des écarts-types des teneurs en

oxygène dissous de l'hypolimnion du bassin est en fonction du jour du calendrier. L'écart-type augmente de façon linéaire en fonction du temps. L'inconstance des variances montre qu'il faudrait faire appel à la régression pondérée pour estimer le taux d'appauvrissement et tester différentes hypothèses relatives à ces taux. Les méthodes utilisées par Dobson et Gilbertson (1971) et Charlton (1979) ne sont donc pas statistiquement efficaces. Afin d'illustrer les effets de l'inégalité des variances sur le calcul des taux d'appauvrissement, nous avons effectué l'analyse à l'aide de trois méthodes différentes: (1) une régression linéaire simple; (2) une régression linéaire pondérée en fonction du nombre de stations d'échantillonnage; et (3) une régression linéaire à pondération empirique (c'est-à-dire tenant compte du nombre de stations et de l'écart-type estimé). Ces trois cas correspondent aux hypothèses suivantes relatives aux erreurs $\varepsilon_i$ du modèle: (1) $\varepsilon_i$ a une distribution normale de moyenne 0 et de variance $\sigma^2$; (2) $\varepsilon_i$ a une distribution normale de moyenne 0 et de variance $\sigma^2/n_i$, où $n_i$ est le nombre de stations de la $i^{ième}$ grappe de l'hypolimnion; et (3) $\varepsilon_i$ a une distribution normale de moyenne 0 et de variance $\sigma_i^2/n_i$. La valeur de $\sigma_i^2$ est inconnue dans le dernier cas, mais elle peut être estimée à l'aide de $s_i^2$ qui est la variance de la concentration d'oxygène dissous dans la $i^{ième}$ grappe de l'hypolimnion. Etant donné que $s_i^2$ est estimé à partir des données, les distributions de probabilité des estimations $b_0$ et $b_1$ de $\alpha_0$ et $\alpha_1$ ne sont pas connues avec précision, mais on peut faire la déduction habituelle si l'on suppose que $s_i^2$ est la vraie valeur de $\sigma_1^2$.

Détection des effets du temps sur la concentration d'oxygène

La question de savoir si les variations d'une année à l'autre de la concentration d'oxygène dissous de l'hypolimnion sont réelles ou simplement aléatoires peut être résolue en vérifiant si les paramètres $\alpha_0$ et $\alpha_1$ du modèle 4.1 sont constants dans le

temps. Brown et coll. (1975) ont décrit plusieurs méthodes graphiques et statistiques permettant de tester cette hypothèse. Dans le cas présent, il faut ajuster le modèle 4.1 aux données en supposant que $\alpha_0$ et $\alpha_1$ ne varient pas d'une année à l'autre, puis calculer les résidus orthogonalisés cumulés et les carrés des résidus cumulés. Ces résidus font ensuite l'objet d'un examen graphique suivi d'un test formel d'indépendance par rapport au temps.

Tests de constance du taux d'appauvrissement

Des tests formels de constance sont effectués dans le but de vérifier les hypothèses suivantes:

(i) H: $\alpha_0$ et $\alpha_1$ sont constants pour l'ensemble des données recueillies;

(ii) $H_1$: $\alpha_0$ est constant pour toutes les années mais $\alpha_1$ peut varier d'une année à l'autre;

(iii) $H_2$: $\alpha_1$ est constant pour toutes les années, mais $\alpha_0$ varie d'une année à l'autre.

Selon la première hypothèse, H, on cherche à vérifier si la concentration initiale d'oxygène au début de la saison de stratification, $\alpha_0$, et le taux d'appauvrissement, $\alpha_1$, demeurent constants pendant la période d'étude. Selon la deuxième hypothèse, $H_1$, on veut vérifier la constance de $\alpha_0$, en admettant la variation du taux d'appauvrissement. Selon la troisième hypothèse, $H_2$, on cherche à vérifier l'absence de variation des taux d'appauvrissement. Tous ces tests s'inscrivent normalement dans une analyse par régression linéaire.

## RESULTATS ET DISCUSSION

Les moyennes et les écarts-types des concentrations d'oxygène de l'hypolimnion déterminés aux mêmes stations la même

année, lors des missions de 1967-1979 et 1967-1980 dans les bassins central et est, sont présentés aux tableaux 43 et 44 respectivement. On y trouvera aussi le nombre d'observations ayant servi au calcul de ces valeurs et la date de mi-mission en jours du calendrier, le $1^{er}$ avril correspondant au jour 0. Les données de ces tableaux ont servi au calcul des taux d'appauvrissement.

Les résultats de l'ajustement du modèle 4.1 selon les trois hypothèses de distribution de $\varepsilon_1$ qualifiées de régression linéaire simple, régression pondérée et régression à poids arbitraires sont présentés aux tableaux 45, 46 et 47. L'ordonnée à l'origine $\alpha_0$ et la pente $\alpha_1$ ont été respectivement estimées à l'aide de $b_0$ et $b_1$. Les valeurs du coefficient de corrélation multiple $R^2$ et le niveau observé de signification de la pente sont aussi présentés pour le cas de la régression linéaire simple. La comparaison des taux d'appauvrissement en oxygène, calculés à l'aide des trois types de régression susmentionnés, révèle que les différences sont négligeables. Cela est prévisible théoriquement, les estimations $b_0$ et $b_1$ de $\alpha_0$ et $\alpha_1$ n'étant pas biaisées, quelle que soit la méthode utilisée. Les tests relatifs à la tendance des taux d'appauvrissement sont cependant, comme nous le verrons plus loin, fortement influencés par la méthode de calcul.

Les taux d'appauvrissement obtenus par régression avec des poids arbitraires, représentés par $b_1$ au tableau 47, ont fait l'objet d'un test de dépendance par rapport au temps basé sur les résidus lors de l'ajustement du modèle 4.1 et aussi sur les résidus orthogonalisés et les carrés des résidus orthogonalisés. On pourra trouver, à la figure 51, le graphique des résidus en fonction de l'ordre séquentiel des observations relatives au bassin central. A la figure 51a, le graphique des résidus normalisés et orthogonalisés montre que ces derniers ne sont pas distribués de façon aléatoire, tandis qu'à la figure 51b, le graphique des résidus cumulés indique un écart systématique par rapport à une situation non aléatoire des

résidus. De plus la figure 51c, où l'on a reporté les résidus orthogonalisés au carré et leurs valeurs prévues, porte à croire à un écart important. Cela vaut aussi pour le bassin est comme en fait foi la figure 52.

On trouvera au tableau 48 les analyses formelles effectuées pour tester les hypothèses de constance des concentrations initiales et des taux de perte d'oxygène dissous dans les bassins central et est. Les résultats sont présentés pour les trois hypothèses différentes relatives à la distribution de $\varepsilon_i$. Il apparait clairement que les valeurs de F diffèrent de façon appréciable. Ainsi, les valeurs de F selon l'hypothèse $H_1$ (taux d'appauvrissement variable) sont de 2.67, 9.11 et 3.80 selon que les données du bassin central sont traitées par régression non pondérée, pondérée ou arbitrairement pondérée. La régression non pondérée se traduit par l'acceptation de l'hypothèse $H_1$, tandis que l'on obtient avec la régression pondérée des différences entre les taux qui sont fortement significatives ($P < 0.01$) et, avec des poids arbitraires, une valeur de $H_1$ significative au niveau de 5 %. Les résultats dépendent donc fortement de l'hypothèse d'égalité des variances. Comme nous l'avons déjà mentionné, l'analyse effectuée ici repose sur une régression à poids arbitraires, nous ne traiterons donc que de ce cas.

Comme on peut le noter à la dernière colonne du tableau 48, les valeurs de F obtenues à partir des hypothèses H, $H_1$ et $H_2$ appliquées au bassin central sont significatives au niveau de 5 %. Les pentes et les ordonnées à l'origine diffèrent donc d'une année à l'autre. Les résultats correspondants obtenus pour le bassin est mènent au rejet des hypothèses H et $H_2$; les taux d'appauvrissement peuvent donc être considérés comme des constantes indépendantes de l'année, mais la concentration initiale d'oxygène au début de la saison de stratification varie d'une année à l'autre.

Ces tests portaient sur des hypothèses de constance dans le temps par rapport à une variabilité générale. La conclusion voulant que ni la concentration d'oxygène dissous initiale, ni le taux d'appauvrissement ne soient constants dans le bassin central et que la concentration d'oxygène dissous initiale ne soit pas constante dans le bassin est, dicte la prochaine étape, soit la détermination de la forme de la dépendance temporelle. Une première étape est présentée à la figure 53 où les taux d'appauvrissement des deux bassins sont reportés en fonction des années. On notera aussi qu'une courbe a été ajustée aux valeurs du bassin central et que l'on a déterminé le taux d'appauvrissement moyen dans le bassin est. Le polynôme quadratique ajusté aux taux d'appauvrissement du bassin central résume la tendance à des taux plus élevés au début et à la fin de la période.

Il serait intéressant de connaître les facteurs ayant pu influer sur les taux d'appauvrissement en oxygène dissous dans le bassin central. Les facteurs qui auraient pu être responsables des faibles taux des années 1972, 1973, 1975 et 1978 sont: (1) les niveaux d'eau élevés en 1972, 1973 et 1975; (2) un hypolimnion très épais en 1975 et 1978; (3) des basses températures dans l'hypolimnion en 1975; (4) un hypolimnion froid au début de la période de stratification d'été en 1975 et 1978. L'important taux d'appauvrissement de 1977 pourrait être dû à une température élevée de l'hypolimnion au début de la période de stratification. El-Shaarawi (Chapitre 5) a traité de l'incorporation de termes représentant certains de ces facteurs dans le modèle de la concentration d'oxygène du bassin central.

Carr (1962) a émis l'hypothèse que le taux d'utilisation de l'oxygène de l'hypolimnion du bassin central était à la hausse. Dobson et Gilbertson (1971) ont accepté cette conclusion générale et calculé une tendance à long terme de 0.075 mg/L par mois et par année pour les années 1930-1970. Ils ont attribué cette tendance à

d'importantes quantités de phytoplancton déposées sur les fonds du lac. Charlton (1979) a, par ailleurs, conclu à l'absence de tendances significatives du taux d'appauvrissement en oxygène dissous suite à une normalisation de ce paramètre visant à tenir compte de facteurs physiques, à savoir la température de l'eau et l'épaisseur de l'hypolimnion. Burns et Rosa (1981) ont, quant à eux, appuyé l'hypothèse d'une tendance, ils ont pris en considération d'autres paramètres physiques dont la température, le mélange vertical et l'apport d'oxygène du bassin est. Nous ne tirerons pas ici une autre conclusion; nous ne traiterons que des étapes qui précèdent cette dernière.

On présente une analyse par regroupement en grappes qui permet de répartir automatiquement les données selon les divers régimes thermiques et spatiaux naturels. Les taux d'appauvrissement en oxygène dissous, calculés pour des stations communes à ces grappes au cours d'une année, sont comparés à la figure 54 aux valeurs obtenues par d'autres. Les valeurs que nous avons calculées se situent généralement entre celles obtenues par d'autres auteurs et, à une exception près, sont inférieures à celles de Herdendorf. L'analyse par grappes représente donc une méthode de remplacement semi-objective et pratique aux méthodes antérieurement utilisées pour le choix des données. L'algorithme de regroupement en grappes présente plusieurs autres caractéristiques intéressantes. Ainsi, il y a réduction de l'importante variabilité spatiale, ce qui permet de rendre plus significatifs les effets liés au temps qui présentent de l'intérêt. D'une façon générale, l'algorithme donne de bons résultats, les moins intéressants étant obtenus avec les données qui correspondent au début de la période de stratification d'été, et sa performance s'accroit à mesure que la thermocline devient plus prononcée.

Les méthodes d'analyse de régression utilisées ici pour estimer la concentration initiale et le taux d'appauvrissement en

oxygène dissous et pour tester les hypothèses quant à leur constance présentent une double importance. Tout d'abord, l'analyse de régression devrait être pondérée de façon à représenter adéquatement les caractéristiques liées à la cueillette des données. Deuxièmement, le modèle décrivant les variations du taux d'appauvrissement peut être formulé de façon à tenir compte de certaines caractéristiques, comme la structure temporelle, lors de l'estimation de ce taux.

Tableau 43. Résumé des concentrations d'oxygène dissous et des profondeurs pour les groupes restreints à des stations communes au cours d'une même année, bassin central, 1967 à 1979.

| Année | Mission | Nombre de points | Jours de calendrier (>1er avril) | Oxygène dissous (mg/L) | | Profondeur (m) | |
|---|---|---|---|---|---|---|---|
| | | | | Moyenne | Ecart type | Moyenne | Ecart type |
| 1967 | 6722103 | 3 | 83.0 | 8.55 | 0.51 | 20.00 | 1.73 |
| | 6722104 | 3 | 98.0 | 7.49 | 1.09 | 21.00 | 1.73 |
| | 6722105 | 3 | 105.0 | 5.88 | 0.60 | 22.00 | 0.00 |
| | 6722107 | 3 | 125.0 | 4.59 | 0.97 | 20.00 | 1.73 |
| | 6722109 | 3 | 146.0 | 1.70 | 0.77 | 19.00 | 0.00 |
| 1968 | 6822102 | 9 | 49.0 | 11.77 | 0.41 | 19.56 | 3.20 |
| | 6822104 | 9 | 77.5 | 8.45 | 0.36 | 20.11 | 3.21 |
| | 6822108 | 9 | 122.0 | 4.24 | 1.22 | 19.67 | 3.27 |
| 1969 | 6922103 | 11 | 62.0 | 10.02 | 0.58 | 20.45 | 2.06 |
| | 6922104 | 11 | 94.0 | 6.60 | 0.56 | 20.81 | 1.53 |
| | 6922105 | 11 | 121.5 | 3.36 | 1.94 | 20.73 | 1.61 |
| | 6922107 | 11 | 147.5 | 0.97 | 0.59 | 21.18 | 2.08 |
| 1970 | 7022103 | 9 | 37.0 | 13.56 | 0.29 | 19.33 | 1.73 |
| | 7022104 | 9 | 64.0 | 10.18 | 0.52 | 20.88 | 1.05 |
| | 7022106 | 9 | 95.5 | 6.45 | 0.36 | 20.55 | 0.88 |
| | 7022109 | 9 | 148.0 | 1.39 | 0.95 | 21.66 | 1.41 |
| 1972 | 7222101 | 12 | 25.5 | 13.25 | 0.38 | 19.58 | 2.07 |
| | 7222102 | 12 | 68.0 | 8.60 | 2.37 | 20.42 | 1.49 |
| | 7222103 | 12 | 88.5 | 9.01 | 0.43 | 20.33 | 1.50 |
| | 7222104 | 12 | 124.5 | 4.50 | 1.14 | 20.25 | 1.65 |
| | 7222106 | 12 | 151.0 | 1.57 | 0.68 | 20.08 | 2.15 |
| 1973 | 7322101 | 28 | 13.0 | 13.36 | 0.39 | 18.80 | 2.23 |
| | 7322103 | 28 | 117.0 | 4.39 | 1.54 | 19.00 | 2.55 |
| | 7322106 | 20 | 150.5 | 2.46 | 2.24 | 19.41 | 2.46 |
| 1975 | 7522102 | 32 | 46.0 | 12.52 | 0.80 | 19.97 | 2.02 |
| | 7522106 | 32 | 87.0 | 9.30 | 1.53 | 20.01 | 2.06 |
| | 7522107 | 32 | 129.0 | 5.80 | 2.32 | 19.92 | 2.26 |
| 1977 | 7722102 | 35 | 63.0 | 10.30 | 0.89 | 20.41 | 2.14 |
| | 7722103 | 35 | 75.5 | 8.85 | 1.33 | 20.23 | 2.08 |
| | 7722105 | 35 | 96.5 | 7.14 | 1.34 | 20.28 | 2.06 |
| | 7722106 | 35 | 117.5 | 4.88 | 1.24 | 20.07 | 2.09 |
| | 7722107 | 35 | 131.5 | 3.62 | 1.91 | 20.11 | 2.11 |
| 1978 | 7822103 | 28 | 60.5 | 11.14 | 1.16 | 19.75 | 2.40 |
| | 7822104 | 28 | 82.5 | 9.45 | 1.91 | 19.99 | 2.31 |
| | 7822106 | 28 | 106.0 | 7.71 | 2.80 | 19.75 | 2.67 |
| | 7822108 | 28 | 124.0 | 6.40 | 2.69 | 19.61 | 2.42 |
| | 7822110 | 28 | 143.0 | 4.69 | 1.98 | 19.76 | 2.47 |
| 1979 | 7922103 | 9 | 46.0 | 12.66 | 0.50 | 21.00 | 1.54 |
| | 7922104 | 9 | 73.0 | 9.66 | 0.57 | 21.22 | 1.35 |
| | 7922106 | 9 | 95.0 | 7.07 | 1.16 | 20.94 | 1.38 |
| | 7922109 | 9 | 115.0 | 4.58 | 1.85 | 21.34 | 1.30 |
| | 7922112 | 9 | 144.5 | 2.48 | 1.73 | 21.44 | 1.16 |

Tableau 44. Résumé des concentrations d'oxygène dissous et des profondeurs pour les groupes restreints à des stationss communes au cours d'une même année, bassin est, 1967 à 1980

| Année | Mission | Nombre de points | Jours de calendrier (>1er avril) | Oxygène dissous (mg/L) | | Profondeur (m) | |
|---|---|---|---|---|---|---|---|
| | | | | Moyenne | Ecart type | Moyenne | Ecart type |
| 1967 | 6722101 | 7 | 61.5 | 12.05 | 0.529 | 38.21 | 11.73 |
| | 6722103 | 7 | 87.5 | 11.44 | 0.35 | 38.57 | 11.28 |
| | 6722105 | 7 | 102.0 | 10.19 | 0.35 | 30.43 | 10.92 |
| | 6722107 | 7 | 122.5 | 9.69 | 0.73 | 38.29 | 13.29 |
| | 6722109 | 7 | 143.5 | 8.57 | 1.05 | 34.57 | 11.53 |
| | 6722111 | 7 | 164.5 | 7.78 | 1.04 | 38.71 | 10.79 |
| 1968 | 6822102 | 4 | 47.0 | 12.70 | 0.23 | 44.00 | 5.94 |
| | 6822104 | 4 | 76.0 | 11.52 | 0.09 | 45.76 | 2.50 |
| | 6822108 | 4 | 120.0 | 9.35 | 0.24 | 49.00 | 8.12 |
| | 6822111 | 4 | 181.0 | 6.62 | 0.41 | 47.75 | 8.22 |
| 1969 | 6922103 | 3 | 60.0 | 12.88 | 0.35 | 52.00 | 10.39 |
| | 6922104 | 3 | 96.0 | 11.11 | 0.44 | 51.00 | 10.44 |
| | 6922105 | 3 | 122.0 | 9.87 | 0.16 | 51.00 | 8.66 |
| | 6922107 | 3 | 150.0 | 9.21 | 0.21 | 52.00 | 9.54 |
| | 6922108 | 3 | 166.0 | 7.98 | 0.72 | 49.68 | 8.96 |
| | 6922110 | 3 | 197.0 | 6.85 | 0.43 | 51.67 | 8.62 |
| 1970 | 7022103 | 3 | 35.0 | 13.64 | 0.06 | 39.00 | 9.54 |
| | 7022104 | 3 | 62.0 | 12.61 | 0.14 | 38.67 | 7.50 |
| | 7022106 | 3 | 93.0 | 11.07 | 0.83 | 41.00 | 9.54 |
| | 7022107 | 3 | 119.0 | 9.60 | 1.24 | 40.00 | 8.72 |
| | 7022109 | 3 | 147.0 | 9.27 | 0.22 | 41.67 | 8.39 |
| | 7022111 | 3 | 175.0 | 7.29 | 0.57 | 41.33 | 8.15 |
| 1972 | 7222101 | 5 | 24.0 | 13.78 | 0.45 | 35.20 | 13.16 |
| | 7222102 | 5 | 67.0 | 12.75 | 0.88 | 35.40 | 13.26 |
| | 7222103 | 5 | 87.0 | 11.53 | 0.75 | 35.20 | 10.57 |
| | 7222104 | 5 | 123.0 | 9.06 | 0.73 | 35.40 | 12.04 |
| | 7222107 | 5 | 181.0 | 6.95 | 1.36 | 36.40 | 11.85 |
| 1973 | 7322101 | 4 | 11.0 | 13.53 | 0.20 | 38.25 | 14.57 |
| | 7322103 | 4 | 115.5 | 9.51 | 0.88 | 38.25 | 13.89 |
| | 7322106 | 4 | 149.5 | 8.07 | 0.71 | 38.75 | 11.70 |
| 1975 | 7522102 | 9 | 46.0 | 13.43 | 0.73 | 44.72 | 9.51 |
| | 7522106 | 9 | 86.5 | 11.41 | 0.56 | 43.61 | 7.37 |
| | 7522107 | 9 | 129.0 | 9.83 | 0.60 | 42.56 | 6.21 |
| | 7522110 | 9 | 191.0 | 8.29 | 0.76 | 42.22 | 7.24 |
| | 7522111 | 9 | 211.0 | 6.72 | 0.72 | 43.47 | 8.13 |
| 1977 | 7722102 | 12 | 60.5 | 12.55 | 0.28 | 40.58 | 6.11 |
| | 7722103 | 13 | 74.0 | 11.81 | 0.21 | 41.15 | 6.50 |
| | 7722105 | 13 | 95.0 | 10.76 | 0.78 | 41.46 | 5.68 |
| | 7722106 | 13 | 116.0 | 9.77 | 0.28 | 40.38 | 5.11 |
| | 7722107 | 13 | 130.0 | 9.08 | 0.51 | 41.15 | 5.58 |
| | 7722109 | 13 | 151.0 | 8.57 | 0.61 | 39.92 | 5.14 |
| | 7722111 | 13 | 172.0 | 5.67 | 0.69 | 36.69 | 5.27 |
| 1978 | 7822103 | 10 | 59.0 | 13.23 | 0.50 | 39.70 | 4.72 |
| | 7822104 | 13 | 80.5 | 11.83 | 0.50 | 40.69 | 5.33 |
| | 7822106 | 13 | 104.5 | 11.29 | 0.40 | 41.19 | 4.82 |
| | 7822108 | 13 | 122.0 | 10.63 | 0.12 | 40.07 | 4.44 |
| | 7822110 | 13 | 141.5 | 9.75 | 0.65 | 38.77 | 4.92 |
| | 7822111 | 13 | 169.0 | 9.03 | 0.54 | 41.35 | 4.72 |
| | 7822114 | 13 | 183.0 | 8.30 | 0.45 | 39.31 | 5.02 |
| 1979 | 7922103 | 4 | 45.0 | 13.44 | 0.14 | 41.50 | 2.89 |
| | 7922104 | 3 | 71.0 | 12.64 | 0.19 | 41.33 | 3.51 |
| | 7922106 | 4 | 93.5 | 11.92 | 0.26 | 41.75 | 2.36 |
| | 7922109 | 4 | 113.5 | 10.73 | 0.43 | 41.75 | 2.87 |
| | 7922112 | 4 | 145.0 | 9.73 | 0.26 | 44.00 | 4.89 |
| | 7922114 | 4 | 178.0 | 8.50 | 0.34 | 46.75 | 9.74 |
| 1980 | 8022101 | 2 | 98.0 | 11.20 | | 42.00 | 4.24 |
| | 8022102 | 2 | 107.0 | 10.75 | 0.04 | 47.50 | 3.54 |
| | 8022104 | 2 | 133.0 | 9.54 | 0.03 | 40.50 | 2.12 |
| | 8022104 | 2 | 142.0 | 10.12 | 0.76 | 41.50 | 4.95 |
| | 8022105 | 2 | 155.0 | 9.00 | 0.45 | 43.00 | 4.24 |

Tableau 45. Résumé des paramètres de régression linéaire simple et des indicateurs de précision de l'analyse de régression des concentrations d'oxygène dissous, bassins central et est

| | Année | Pente estimée* $b_1$ | Ordonnée à l'origine estimée $b_0$ | Coefficient de corrélation multiple $R^2$ | Niveau de signification de la pente $\alpha$ |
|---|---|---|---|---|---|
| Bassin central | 1967 | 0.1078 | 17.6529 | 0.9790 | 0.005 |
| | 1968 | 0.1023 | 16.6277 | 0.9966 | 0.05 |
| | 1969 | 0.1071 | 16.6183 | 0.9981 | 0.005 |
| | 1970 | 0.1092 | 17.3028 | 0.9957 | 0.005 |
| | 1972 | 0.0903 | 15.6452 | 0.9644 | 0.005 |
| | 1973 | 0.0809 | 14.3037 | 0.9952 | 0.05 |
| | 1975 | 0.0810 | 16.2783 | 0.9997 | 0.025 |
| | 1977 | 0.0967 | 16.3164 | 0.9979 | 0.005 |
| | 1978 | 0.0772 | 15.8400 | 0.9986 | 0.005 |
| | 1979 | 0.1059 | 17.3228 | 0.9905 | 0.005 |
| Bassin est | 1967 | 0.0430 | 14.8409 | 0.9819 | 0.005 |
| | 1968 | 0.0459 | 14.9075 | 0.9993 | 0.005 |
| | 1969 | 0.0435 | 15.3807 | 0.9905 | 0.005 |
| | 1970 | 0.0441 | 15.2171 | 0.9821 | 0.005 |
| | 1972 | 0.0464 | 15.2841 | 0.9732 | 0.005 |
| | 1973 | 0.0392 | 13.9762 | 0.9996 | 0.025 |
| | 1975 | 0.0374 | 14.9026 | 0.9797 | 0.005 |
| | 1977 | 0.0553 | 16.0469 | 0.9486 | 0.005 |
| | 1978 | 0.0373 | 15.1644 | 0.9867 | 0.005 |
| | 1979 | 0.0381 | 15.2578 | 0.9936 | 0.005 |
| | 1980 | 0.0345 | 14.4974 | 0.8602 | 0.025 |

* Toutes les pentes sont négatives.

Tableau 46. Estimation des pentes et des ordonnées à l'origine obtenues par régression avec pondération

| | Bassin central | | Bassin est | |
|---|---|---|---|---|
| Année | Pente* | Ordonnée à l'origine | Pente* | Ordonnée à l'origine |
| 1967 | 0.1077 | 17.645 | 0.0430 | 14.838 |
| 1968 | 0.1024 | 16.636 | 0.0458 | 14.905 |
| 1969 | 0.1071 | 16.618 | 0.0425 | 15.240 |
| 1970 | 0.1092 | 17.303 | 0.0441 | 15.217 |
| 1972 | 0.0903 | 15.645 | 0.0464 | 15.284 |
| 1973 | 0.0814 | 14.321 | 0.0392 | 13.976 |
| 1975 | 0.0810 | 16.278 | 0.0374 | 14.903 |
| 1977 | 0.0967 | 16.317 | 0.0553 | 16.056 |
| 1978 | 0.0772 | 15.840 | 0.0370 | 15.110 |
| 1979 | 0.1059 | 17.323 | 0.0380 | 15.246 |
| 1980 | -- | -- | 0.0345 | 14.498 |

* Toutes les pentes sont négatives.

Tableau 47. Estimation des pentes et des ordonnées à l'origine obtenues par régression avec poids arbitraires

| | Bassin central | | Bassin est | |
|---|---|---|---|---|
| Année | Pente* | Ordonnée à l'origine | Pente* | Ordonnée à l'origine |
| 1967 | 0.1070 | 17.415 | 0.0444 | 14.993 |
| 1968 | 0.1089 | 17.023 | 0.0464 | 15.011 |
| 1969 | 0.1060 | 16.572 | 0.0410 | 15.072 |
| 1970 | 0.1165 | 17.775 | 0.0402 | 15.057 |
| 1972 | 0.0875 | 15.817 | 0.0446 | 15.010 |
| 1973 | 0.0836 | 14.442 | 0.0392 | 13.962 |
| 1975 | 0.0802 | 16.219 | 0.0373 | 14.824 |
| 1977 | 0.0975 | 16.401 | 0.0520 | 15.705 |
| 1978 | 0.0773 | 15.828 | 0.0371 | 15.148 |
| 1979 | 0.1090 | 17.415 | 0.0372 | 15.175 |
| 1980 | -- | -- | 0.0469 | 15.787 |

* Toutes les pentes sont négatives.

Tableau 48. Résultats des tests d'égalité des pentes et des ordonnées à l'origine - Modèle 4.1

| Bassins | Modèles | Degrés de liberté | Valeurs de F non pondérées | Valeurs de F pondérées | Valeurs de F avec poids arbitraires |
|---|---|---|---|---|---|
| Central | Pente constante et ordonnées à l'origine différentes | 9,22 | 3.62* | 4.05* | 4.47* |
| | Pentes différentes et ordonnée à l'origine constante | 9,22 | 2.67 | 9.11** | 3.80* |
| | Pente et ordonnée à l'origine constantes | 18,22 | 6.58** | 7.77** | 11.97** |
| Est | Pente constante et ordonnée à l'origine différentes | 10,38 | 2.59* | 3.25* | 4.63** |
| | Pentes différentes et ordonnée à l'origine constante | 10,38 | 1.55 | 1.30 | 1.38 |
| | Pente et ordonnée à l'origine constantes | 20,38 | 6.59** | 6.74** | 15.15** |

* Significatif au niveau de 5 %.
** Significatif au niveau de 1 %.

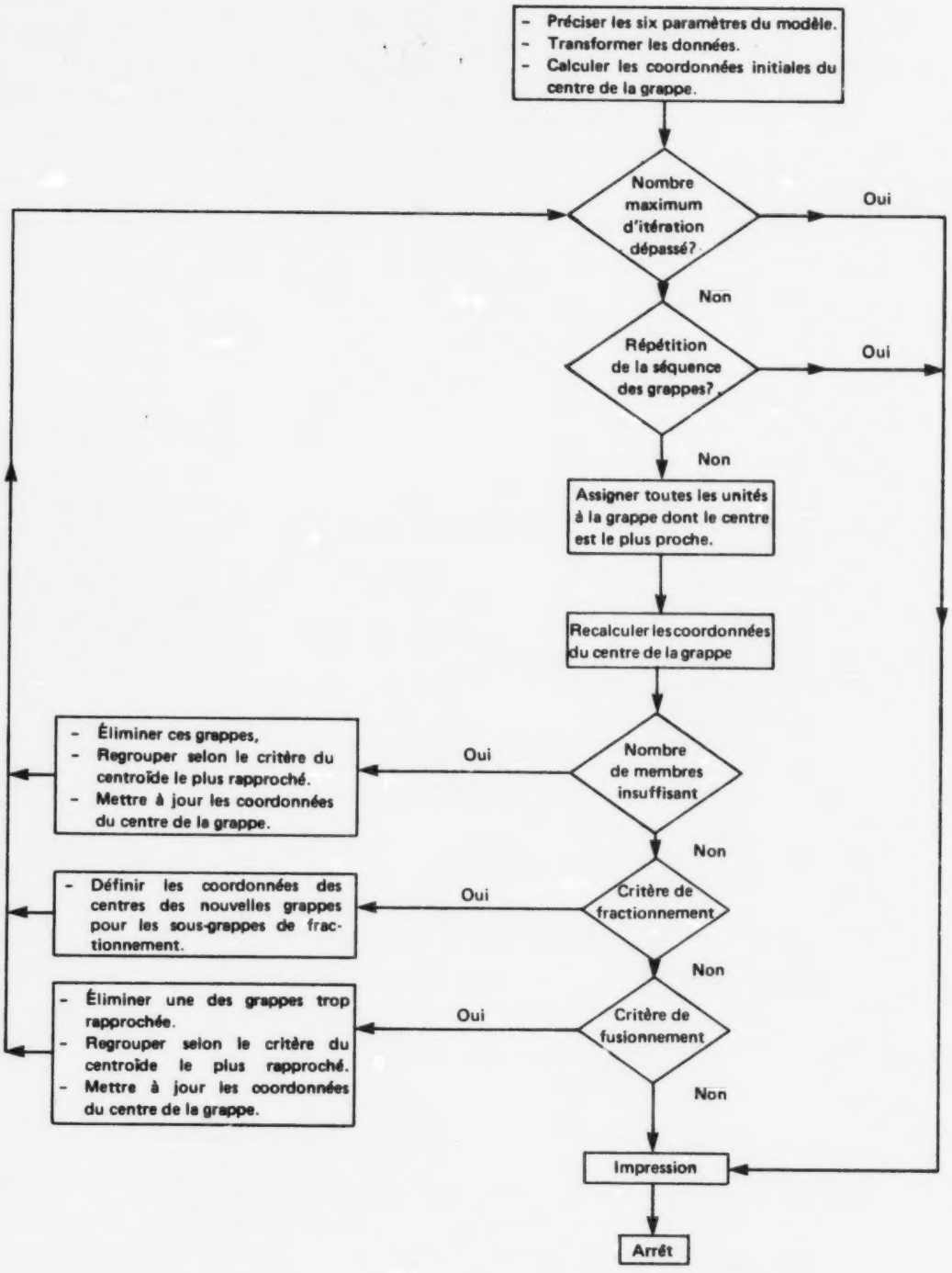



**Figure 48. Organigramme de l'algorithme de regroupement en grappes.**

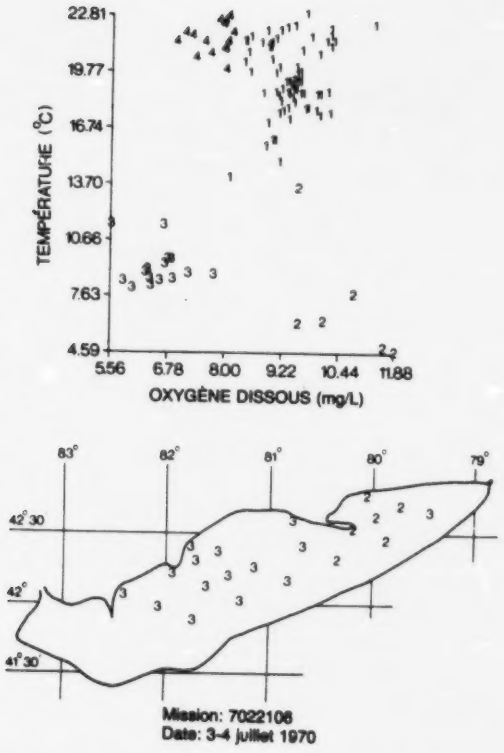



**Figure 49. Exemple de tracés générés par le programme de regroupement, mission 7022106.**

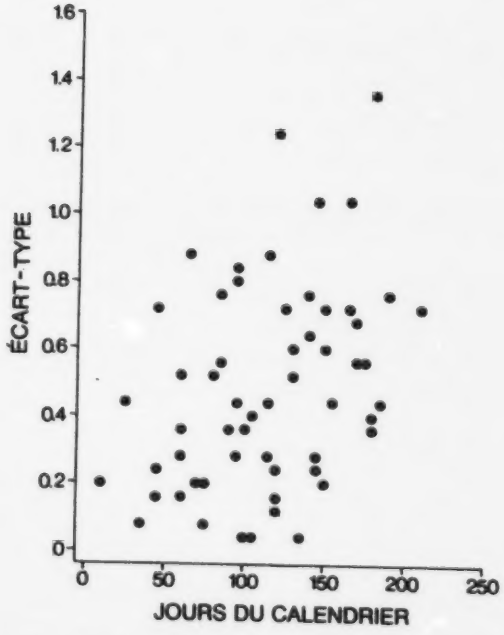



**Figure 50. Écart-type de la concentration d'oxygène dans l'hypolimnion pour les grappes du bassin est en fonction du jour du calendrier.**

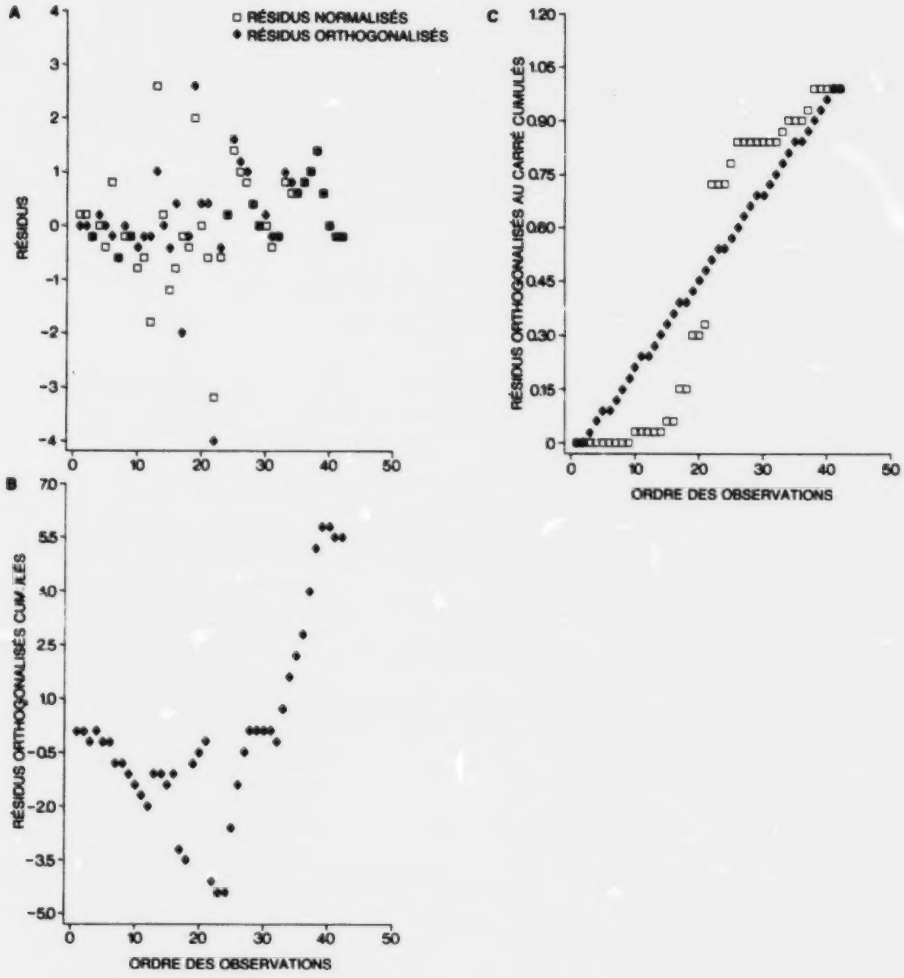



**Figure 51.** Reports des résidus dans l'ordre des observations au bassin central : (a) résidus normalisés et orthogonalisés; (b) résidus cumulés; (c) résidus orthogonalisés au carré.

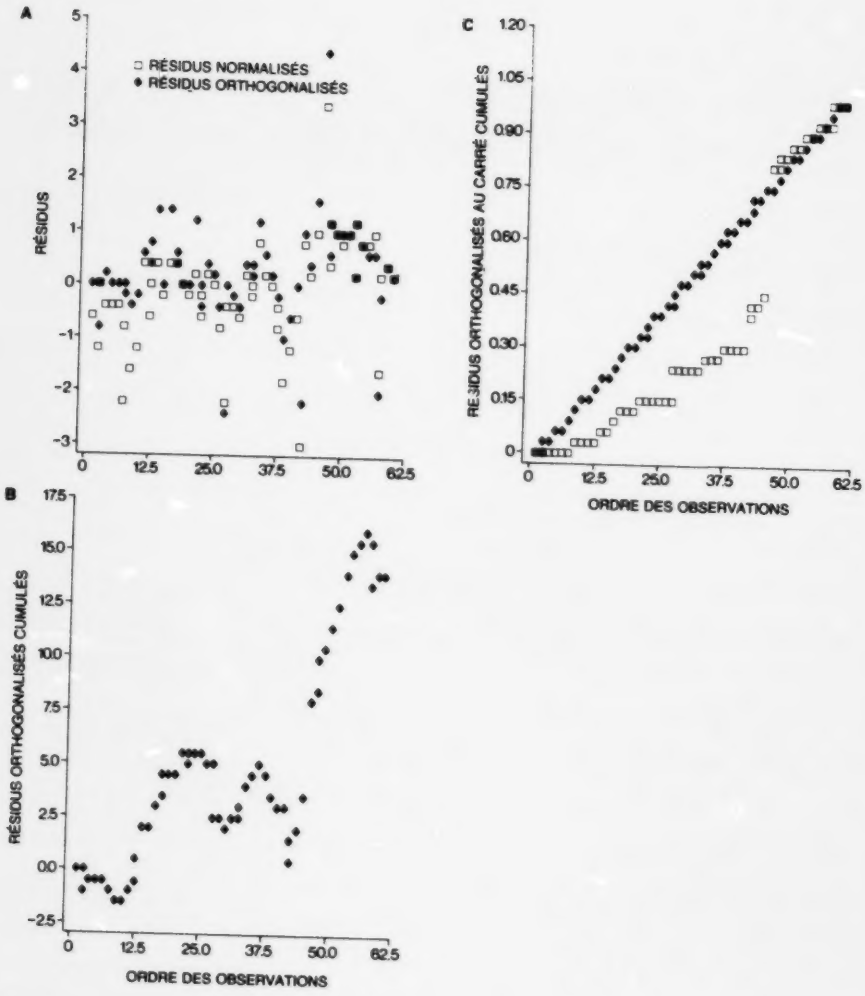



**Figure 52.** Reports des résidus dans l'ordre des observations au bassin est : (a) résidus normalisés et orthogonalisés; (b) résidus cumulés. (c) résidus orthogonalisés au carré.

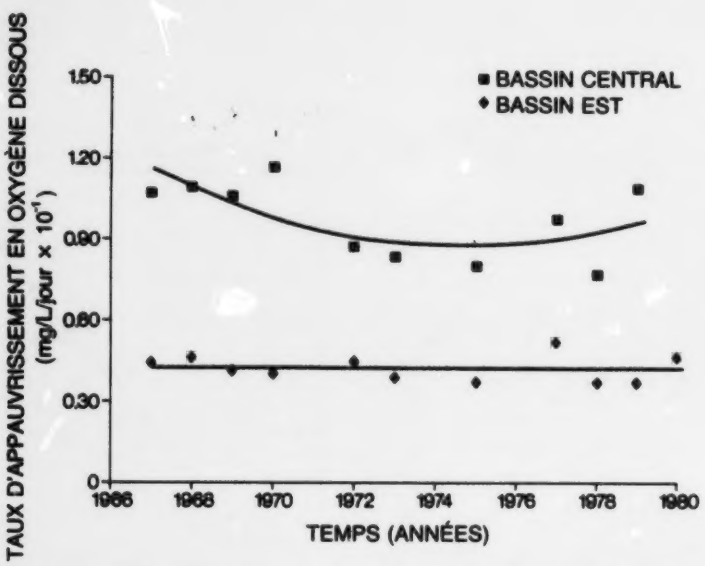



**Figure 53. Taux d'appauvrissement en oxygène dissous en fonction de l'année et courbes ajustées, bassins central et est.**

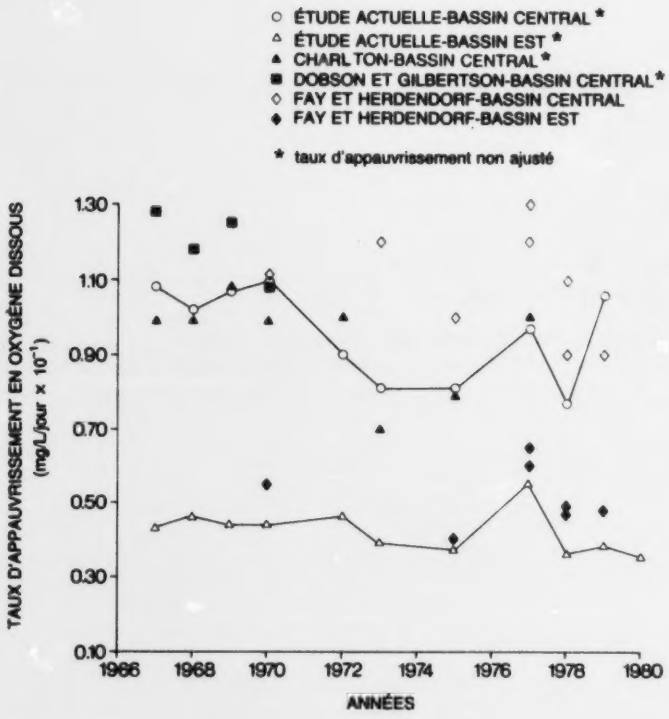



**Figure 54. Comparaison des taux d'appauvrissement en oxygène dissous non corrigés tels que calculés par divers auteurs.**

# Modèle statistique de l'oxygène dissous dans le bassin central du lac Érié

*par A.H. El-Shaarawi*

## INTRODUCTION

Anderson et coll. (chapitre 4) ont démontré que la concentration d'oxygène dans l'hypolimnion du bassin central du lac Erié a, de 1967 à 1979, fluctué de façon statistiquement significative d'une année à l'autre, quel que soit le moment choisi de la période de stratification d'été. La concentration moyenne d'oxygène dans un bassin hypolimnétique diminuant de façon linéaire au travers de la stratification, les différences d'une année à l'autre peuvent être divisées en deux composantes corrélées, mais distinctes. La première est due aux variations de l'ordonnée à l'origine de la droite, et la deuxième, aux variations de la pente de cette droite. Il s'avère plus pratique de considérer l'ordonnée à l'origine de la pente comme représentant, respectivement, la concentration d'oxygène au début de la période de stratification et le taux d'appauvrissement. Au chapitre 4, on n'a pas tenté d'expliquer les variations de la teneur initiale ou du taux d'appauvrissement en oxygène, mais les auteurs ont résumé la tendance du taux d'appauvrissement par une équation quadratique. Dobson et Gilbertson (1971) en sont venus à la conclusion que le taux d'appauvrissement de 1970 était supérieur de plus du double au taux estimé pour 1929. Ils ont attribué cet accroissement à des augmentations de la production phytoplanctonique suite à des apports accrus de matières nutritives. Charlton (1980a,b) a conclu que les taux d'appauvrissement ne présentaient pas de tendance à long terme

et que les différences apparentes sont en grande partie liées aux variations d'épaisseur et de température de l'hypolimnion. Barica (1982) a mentionné que les résultats de Charlton avaient récemment été mis en doute par F. Rosa et N.M. Burns. Il faut mentionner que, à l'exception de Anderson et coll. (chapitre 4), aucun des auteurs cités plus haut n'a effectué ou présenté de test statistique portant sur la tendance des taux d'appauvrissement.

Thienemann (1915) a distingué cinq facteurs ayant une incidence sur le taux d'appauvrissement: (1) la saison; (2) la situation du lac par rapport aux vents dominants; (3) le volume de l'hypolimnion et le rapport entre les eaux sus-jacentes et sous-jacentes à la thermocline; (4) la température des eaux de l'hypolimnion; et (5) la quantité de matières organiques transportées vers les eaux de fond. Il apparaît clairement que nous ne pouvons agir que sur le dernier facteur,surtout en luttant contre la pollution industrielle et agricole. Il est reconnu que le phosphore est un facteur limitatif de la productivité végétale dans les Grands lacs. La Commission mixte internationale (CMI) a, dans le but de lutter contre la pollution, fixé des objectifs pour la charge en phosphore dans chacun des Grands lacs. Slater et Bangay (1980) ont signalé que, dans le lac Erié, la charge par source ponctuelle avait été réduite d'environ 10 000 tonnes métriques par an en 1972/1973 à 5700 tonnes métriques par an en 1977; il est prévu que cette valeur sera de 2100 tonnes métriques par an quand toutes les mesures de lutte permettant de respecter la norme de 1 mg/L de phosphore dans les effluents auront été mises en place. Il est important d'étudier les effets de ces étapes de lutte antipollution sur la concentration d'oxygène dissous dans l'hypolimnion du lac Erié.

Dans le présent chapitre, on élabore un modèle statistique des concentrations d'oxygène dissous dans l'hypolimnion. Le niveau d'eau du lac, la température de l'hypolimnion et les concentrations moyennes de phosphore total sont les variables explicatives. La

probabilité d'anoxie dans le bassin central est estimée en fonction de la teneur en phosphore, de la température et du niveau d'eau. De plus, le modèle sert à prévoir les teneurs en phosphore correspondant à des taux d'appauvrissement et à des niveaux d'eau donnés. On traite aussi de l'utilisation du modèle dans la réglementation et l'établissement de normes.

## ELABORATION DU MODELE

Les résultats obtenus par Anderson et coll. (chapitre 4) indiquent que la concentration initiale d'oxygène dissous dans l'hypolimnion, $\alpha_0$, et le taux d'appauvrissement, $\alpha_1$, varient d'une année à l'autre. Comme nous l'avons indiqué plus haut, plusieurs facteurs sont responsables de cette variation. Nous traiterons ici des effets du niveau d'eau, de la température de l'eau de l'hypolimnion et de la concentration moyenne annuelle de phosphore total et utiliserons ces paramètres pour la modélisation des concentrations d'oxygène dissous dans le bassin central du lac Erié.

On trouvera aux figures 55 et 56 les estimations de l'ordonnée à l'origine, $b_0$, et du taux d'appauvrissement $b_1$, en fonction du niveau de l'eau du lac Erié. Les valeurs $b_0$ et $b_1$ sont les estimations respectives de la concentration initiale d'oxygène dissous au début de la période de stratification et du taux d'appauvrissement en oxygène et, partant, les estimations des paramètres $\alpha_0$ et $\alpha_1$ du modèle 5.1 de Anderson et coll. pour les bassins central et est. Ce modèle s'exprime sous la forme

$$y_i = \alpha_0 + \alpha_1 x_i + \epsilon_i \quad i = 1,2...NM \qquad (5.1)$$

où NC est le nombre de missions effectuées au cours de l'année étudiée, $y_i$ est la concentration moyenne de $O_2$ (mg/L), $x_i$ est la date du milieu de la mission (jour du calendrier débutant en avril) et $\epsilon_i$ est une variable aléatoire de moyenne 0.

Les valeurs du bassin est n'indiquent pas l'existence d'une relation entre le niveau de l'eau et $b_0$ ou $b_1$, mais les valeurs de $b_0$ et $b_1$ du bassin central présentent une corrélation fortement négative avec le niveau de l'eau. Comme les valeurs de $b_0$ semblent varier de façon linéaire avec le niveau de l'eau, on a décidé de vérifier si les fluctuations de l'ordonnée à l'origine $b_0$ pouvaient faire l'objet d'une modélisation utilisant le niveau de l'eau comme covariable. Le modèle 5.1 devient donc

$$y_{ij} = \beta_0 + \beta_1 w_j + \gamma_j x_{ij} + \epsilon_{ij} \quad \begin{matrix} i = 1,2...NM_j \\ j = 1,2...m \end{matrix} \qquad (5.2)$$

où $NM_j$ est le nombre de missions au cours de la $j^{ième}$ année, m est le nombre d'années, $y_{ij}$ est la concentration moyenne d'oxygène dissous lors de la $i^{ième}$ mission de la $j^{ième}$ année, $x_{ij}$ est la date médiane de la mission en jours du calendrier lors de la $i^{ième}$ mission de la $j^{ième}$ année, $w_j$ est le niveau de l'eau lors de la $j^{ième}$ année, $\epsilon_{ij}$ est une variable aléatoire normale de moyenne 0 et de variance estimée $S_{ij}/N_{ij}$, où $S_{ij}$ est l'écart type de la concentration d'oxygène dissous au cours de la $j^{ième}$ année et $N_{ij}$ est le nombre correspondant de stations d'échantillonnage, et $\beta_0$, $\beta_1, \gamma_1 \cdots \gamma_m$ sont les paramètres inconnus du modèle. Les paramètres $\gamma_1 \cdots \gamma_m$ sont les taux d'appauvrissement en oxygène ajustés; m est égal à 10 dans le présent cas.

Le modèle 5.2 a été ajusté aux données du bassin central. Les résultats obtenus indiquent que l'inclusion du niveau de l'eau permet d'expliquer complètement la variabilité d'une année à l'autre de l'ordonnée à l'origine du modèle 5.2. La statistique

$$LF = 22\{RES(2) - RES(1)\}/\{8 \times RES(1)\}$$

est une mesure du manque d'ajustement du modèle 5.2, comparativement au modèle 5.1; RES (1) et RES (2) sont les sommes des carrés

résiduelles respectives des modèles 5.1 et 5.2. La statistique LF a une distribution F à 8 et 22 degrés de liberté. La valeur observée de LF est 0.95 et n'est pas significative. Les estimations des paramètres du modèle 5.2 sont présentées au tableau 49. Les quantités $\hat{\gamma}_1$, $\hat{\gamma}_2 \cdots \hat{\gamma}_{10}$ sont des estimations des taux d'appauvrissement en oxygène dissous $\gamma_1$, $\gamma_2 \cdots \gamma_{10}$ pour les dix années de l'étude. On note au tableau que les valeurs de $\hat{\gamma}_1$, $\hat{\gamma}_2 \cdots \hat{\gamma}_{10}$ varient de façon appréciable d'une année à l'autre. L'égalité des $\gamma_i$ a été testée par la statistique F et les résultats obtenus montrent que les différences entre les taux d'appauvrissement sont fortement significatives ($P < 0.01$ et $F_{9,30} = 12.475$). Les estimations $\hat{\gamma}_1 \cdots \hat{\gamma}_{10}$ ont été reportées à la figure 57 en fonction des niveaux d'eau correspondants. On note entre elles une forte corrélation négative (presque linéaire). Les valeurs de $\gamma_j$ sont donc modélisées par

$$\gamma_j = \theta_0 + \theta_1 (w_j - \bar{w}) \tag{5.3}$$

où $\theta_0$ et $\theta_1$ sont des constantes, $w_j$ est le niveau d'eau en pieds au cours de la $j^{ième}$ année et $\bar{w}$ est la moyenne des niveaux d'eau annuels au cours de la période d'étude. Les valeurs de $\gamma_j$ du modèle 5.2 sont remplacées par le membre de droite de l'expression 5.3. Nous obtenons ainsi le nouveau modèle

$$\begin{aligned} y_{ij} = {} & \beta_0 + \beta_1 (w_j - \bar{w}) + \theta_0 (x_{ij} - \bar{x}) + \theta_1 (w_j - \bar{w}) \\ & (x_{ij} - \bar{x}) + \varepsilon_{ij}, \qquad i = 1,2 \ldots NM_j \\ & \qquad\qquad\qquad\quad j = 1,2 \ldots m \end{aligned} \tag{5.4}$$

où $\bar{x}$ est la date moyenne de la mission en jours du calendrier pour toutes les années. L'équation de régression estimée est:

$$\begin{aligned} y_{ij} = {} & 7.0043 + 0.4147 (w_j - \bar{w}) \quad -0.1007 (x_{ij} - \bar{x}) \\ & + 0.0233 (x_{ij} - \bar{x})(w_j - \bar{w}) \end{aligned}$$

Le test de manque d'ajustement pour comparer le modèle 5.2 au modèle 5.4 indique que le modèle 5.4 ($P < 0.01$, $F_{8,30} = 6.80$) ne convient pas car il ne permet pas d'expliquer totalement la variation des taux d'appauvrissement.

D'autres tentatives visant à expliquer le reste de la variation des taux d'appauvrissement ont consisté à introduire la température de l'hypolimnion. El-Shaarawi (1984a) a noté que la température de l'hypolimnion fluctue d'une année à l'autre et, même, parfois au cours d'une même année. De fait, il a représenté la température par une droite de régression en fonction du temps. On obtient ainsi une valeur plus représentative de la température moyenne de l'hypolimnion quand on choisit sur la droite de régression le jour 200 du calendrier (notons qu'on situe la période de stratification entre les jours 150 et 250). Si l'on exprime par $T_j$ la température de l'hypolimnion au jour 200 de la $j^{ième}$ année et par $\bar{T}$ la température moyenne pour toutes les années, il faut alors modifier le modèle 5.4 ainsi:

$$y_{ij} = \beta_0 + \beta_1 (w_j - \bar{w}) + \theta_1 (x_{ij} - \bar{x}) + \theta_1 (w_j - \bar{w}) (x_{ij} - \bar{x}) + \alpha_1 (T_j - \bar{T}) + \varepsilon_{ij} \qquad (5.5)$$

La valeur estimée de la concentration d'oxygène dissous est:

$$\hat{y}_{ij} = 7.0921 + 0.1522 (w_j - \bar{w}) - 0.10085 (x_{ij} - \bar{x}) + 0.020881 (x_{ij} - \bar{x}) (w_j - \bar{w}) - 0.28821 (T_j - \bar{T}) \qquad (5.6)$$

La statistique de manque d'ajustement $F = 4.64$ s'avère significative lorsque comparée à la valeur de la fonction de distribution F à 7 et 30 degrés de liberté. D'autres tracés et l'examen des résidus ont montré que le modèle 5.6 devrait comporter un terme d'interaction

entre la température de l'hypolimnion et le niveau de l'eau. On a alors obtenu la forme estimée du modèle pour l'oxygène dissous:

$$\hat{y}_{ij} = 7.011 + 0.1128\,(w_j - \bar{w}) - 0.10015\,(x_{ij} - \bar{x}) + 0.017287\,(x_{ij} - \bar{x})(w_j - \bar{w}) - 0.407859\,(T_j - \bar{T}) - 0.53079\,(T_j - \bar{T})(w_j - \bar{w}) \qquad (5.7)$$

La statistique de manque d'ajustement F est égale à 2.704, valeur significative au niveau de 5 % lorsque comparée à la valeur de la fonction de distribution F à 6 et 30 degrés de liberté. On a décidé de vérifier si l'introduction du terme d'interaction entre la température de l'hypolimnion et le niveau de l'eau au modèle 5.7 se traduisait par une diminution de l'effet principal du niveau de l'eau (le terme exclusif du niveau d'eau) dans le modèle. Le test F a fait accepter l'hypothèse voulant que le terme du niveau d'eau est inutile dans le modèle 5.7. La valeur observée de F est 0.3613 et n'est pas significative au niveau de 5 %. La concentration d'oxygène dissous en fonction du niveau du lac et de la température de l'hypolimnion est:

$$\hat{y}_{ij} = 7.014 - 0.10005\,(x_{ij} - \bar{x}) + 0.015896\,(w_j - \bar{w})(x_{ij} - \bar{x}) - 0.42064(T_j - \bar{T}) - 0.535174\,(T_j - \bar{T})(w_j - \bar{w}) \qquad (5.8)$$

La statistique de manque d'ajustement du modèle T a une valeur significative au niveau de 5 % de 2.58.

Cette analyse montre que le niveau de l'eau et la température de l'hypolimnion ne peuvent expliquer complètement toute la variation non aléatoire de la concentration d'oxygène de l'hypolimnion. On tentera donc, dans le reste de la section, de déterminer l'effet du processus d'eutrophisation sur la concentration

d'oxygène. Soit $P_j$ (j = 1,2...10) la concentration moyenne du phosphore total (PT) de l'épilimnion du bassin central du lac Erié au cours de la $j^{ième}$ année. L'inclusion du PT se traduit par la modification du modèle 5.8 en

$$\hat{y}_{ij} = 7.063 - 0.10073\,(x_{ij} - \bar{x}) + 0.018059\,(w_j - \bar{w})(x_{ij} - \bar{x}) \qquad (5.9)$$
$$- 0.3436(T_j - \bar{T}) - 0.4677\,(T_j - \bar{T})(w_j - \bar{w}) - 51.1339(P_j - \bar{P})$$

où $\bar{P}$ est la moyenne arithmétique de $P_1 \ldots P_{10}$. La statistique de manque d'ajustement du modèle 5.9 est 1.84 et n'est pas significative au niveau de 5 % lorsque comparée à la distribution F à 6 et 30 degrés de liberté. Même si le modèle 5.9 permet d'expliquer la plus grande partie de la variation significative des données, on a été d'avis que le PT ne devrait pas influer uniquement sur la concentration initiale d'oxygène de l'hypolimnion au début de la saison de stratification, mais qu'il devrait aussi influer sur le taux d'appauvrissement en oxygène. Les sommes des carrés résiduelles des modèles 5.2 et 5.9 sont, respectivement, de 165.572 et 226.428; on voit donc que la différence entre les sommes des carrés résiduelles des deux modèles est importante bien que non statistiquement significative. On a donc ajouté un autre terme au modèle 5.9 afin de tenir compte de l'interaction entre le temps, en jours du calendrier, et le PT. Cela a permis de réduire la différence de 226.428 à 194.654; ce qui est très appréciable. Le nouveau modèle ainsi obtenu est:

$$\hat{y}_{ij} = 7.1037 - 0.09778\,(x_{ij} - \bar{x}) + 0.01448\,(x_{ij} - \bar{x})(w_j - \bar{w})$$
$$- 0.36668(T_j - \bar{T}) - 0.4978(T_j - \bar{T})(w_j - \bar{w}) - 71.4513(P_j - \bar{P}) \qquad (5.10)$$
$$- 1.4575\,(x_{ij} - \bar{x})(P_j - \bar{P})$$

et la statistique du manque d'ajustement prend la valeur non significative de 1.054 qui est voisine de un (valeur théorique prévue). Le modèle 5.10 a donc été choisi comme modèle définitif de la concentration d'oxygène dans l'hypolimnion du bassin central. Les valeurs de $\bar{x}$, $\bar{T}$, $\bar{w}$ et $\bar{p}$ sont respectivement 96.393, 10.393, 571.881 et 0.0180. Si l'on suppose que le modèle 5.10 est vraiment représentatif, la dépendance du taux d'appauvrissement par rapport au PT est significative au niveau de 5 % ($t_{35}$ = 2.391) et la dépendance de la concentration d'oxygène par rapport à l'interaction entre la température et le niveau de l'eau est fortement significative ($t_{35}$ = 3.52). De plus, l'effet du PT sur la concentration d'oxygène est extrêmement significatif ($F_{2.35}$ = 6.333). L'inclusion du PT dans le modèle s'est traduite par une réduction très importante de la somme des carrés résiduelle qui est passée de 265.09 à 194.654. On trouvera à la figure 58 le report des concentrations d'oxygène observées ainsi que les droites de régression estimées en fonction des jours du calendrier, le premier avril étant le jour 0. Le modèle est raisonnablement bien ajusté aux données.

A noter: le modèle 5.10 montre que, pendant la période de 1967 à 1979 et pour les seules données recueillies par le C.C.E.I., les fluctuations de la concentration d'oxygène peuvent s'expliquer par le niveau de l'eau, la température de l'hypolimnion et la teneur en PT. Ce modèle sera utilisé à partir de maintenant pour: (i) prévoir le taux d'appauvrissement en fonction du niveau de l'eau et de la teneur en PT; (ii) décrire la tendance de séries chronologiques; (iii) estimer la teneur en PT à partir du taux d'appauvrissement, du niveau de l'eau et de la température de l'hypolimnion; et (iv) estimer la probabilité d'anoxie correspondant à divers niveaux d'eau, températures de l'hypolimnion, teneurs en phosphore total et longueurs de la période de stratification.

## PREVISION DU TAUX D'APPAUVRISSEMENT ET TENDANCE DES SERIES CHRONOLOGIQUES

On note au modèle 5.10 que le signe négatif du coefficient de $(x_{jj} - \bar{x})$ est l'estimation du taux d'appauvrissement au cours de la $j^{ième}$ année; on obtient donc:

$$D_j = 0.09778 - 0.01448\ (w_j - \bar{w}) + 1.4575\ (P_j - \bar{P}) \qquad (5.11)$$

Le taux d'appauvrissement décroît à mesure que le niveau de l'eau augmente et croît avec la teneur en phosphore total. On a reporté à la figure 59 les taux d'appauvrissement en fonction du niveau de l'eau pour diverses teneurs en PT. Ainsi, un taux d'appauvrissement de 0.10 mg/L par jour dans l'hypolimnion du bassin central correspond à un niveau d'environ 569.5 pieds si la teneur en PT est très faible (presque 0) et de 571 pieds si la teneur en PT est de 0.01. On présente aussi à la figure 59 les taux d'appauvrissement des années antérieures à 1967 tels que calculés par Charlton (1979); ceux des années 1967 à 1979 sont les valeurs signalées par Anderson et coll. au chapitre 4. On peut y noter un accroissement progressif des teneurs en PT de 1948 à 1970, suivi d'un déclin de 1972 à 1978. La valeur de 1979 indique une augmentation du PT. Cette déduction est clairement confirmée par le fait que les taux d'appauvrissement de 1962, 1968, 1969, 1970 et 1979 se situent entre les valeurs correspondant aux teneurs en PT de 0.02 et de 0.03 tandis que les taux de 1948, 1949, 1950, 1951, 1952, 1953 et 1963 sont inférieurs à ceux correspondant à la teneur en PT de 0.01. Cette déduction est basée sur des valeurs moyennes (les fluctuations statistiques des taux d'appauvrissement estimés sont ignorées) dont on traite ci-après.

Les intervalles de confiance pour estimer les taux d'appauvrissement et tester diverses hypothèses relatives à leurs valeurs sont établis de la façon suivante. Soit C la matrice des variances-covariances des paramètres du modèle 5.10, alors la

sous-matrice $C_1$ de C, qui contient les éléments de C correspondant aux coefficients de $(x_{ij} - \bar{x})$, $(w_j - \bar{w})(x_{ij} - \bar{x})$ et $(x_{ij} - \bar{x})(P_j - \bar{P})$, est la matrice des variances-covariances des coefficients. Ainsi, pour tout niveau d'eau w' et toute teneur en phosphore total P', la variance de D est:

$$\text{var}(D) = \underline{a}' C_1 \underline{a} \tag{5.12}$$

où $\underline{a}$ est le vecteur colonne (1 w'-w P'-P) et $\underline{a}'$ est la transposée de $\underline{a}$. L'intervalle de confiance à 1-2α de D est:

$$D \pm t_{d,\alpha} \sqrt{\underline{a}' C_1 \underline{a}} \tag{5.13}$$

où $t_{d,\alpha}$ est la valeur, tirée des tables, de la distribution t à d degrés de liberté pour un niveau de signification de 2α.

Cet intervalle de confiance pour D peut être transformé en intervalle de confiance pour chacune des deux variables indépendantes si l'on connait le taux d'appauvrissement et la valeur de l'autre variable indépendante. Ainsi, il est possible d'obtenir un intervalle de confiance pour la teneur en PT si l'on connait le taux d'appauvrissement, le niveau de l'eau et la température de l'hypolimnion. Soit $D_0$ le taux d'appauvrissement quand le niveau de l'eau est égal à $w_0$. L'estimation $P_0$ de PT peut être obtenue pour une valeur donnée de $w_0$ du niveau d'eau en mettant en équation $D_j$ et $D_0$ et en déterminant la valeur de $P_j$ dans l'équation 5.11. Cette valeur de $P_j$ est l'estimation recherchée de PT. L'intervalle de confiance 1-2α de $P_0$ consiste en toutes les valeurs de $P_j$ qui satisfont l'inéquation

$$\frac{(D_0 - D_j)^2}{\underline{a}' C_1 \underline{a}} \leq t^2_{d,\alpha} \tag{5.14}$$

où, dans l'expression de $D_j$ et de $\underline{a}$, on pose $w_j$ égal à $w_0$.

A partir des taux signalés par Charlton et des observations du niveau d'eau, on a estimé l'intervalle de confiance à 95 % des teneurs en PT pour les années étudiées par Charlton (1979). Les résultats sont présentés à la figure 60. Charlton a signalé deux taux d'appauvrissement pour certaines années, auxquels cas on a présenté deux intervalles de confiance qui sont représentés par des tirets, les autres années étant représentées par des traits continus. Comme on peut le voir à la figure, les intervalles de confiance sont généralement décalés vers la droite, ce qui indique une augmentation de la teneur en phosphore total. L'accroissement le plus important est probablement survenu après 1952.

## PREVISION DE L'ANOXIE DE L'HYPOLIMNION

Le modèle 5.10 peut renseigner sur les rôles conjugués de la température de l'hypolimnion, du niveau de l'eau et de la teneur en phosphore total dans l'anoxie du lac. Un examen détaillé du modèle 5.10 permet de découvrir (i) que le taux d'appauvrissement décroît à mesure qu'augmente le niveau de l'eau, mais qu'il croît avec la teneur en PT et (ii) que la concentration initiale d'oxygène de l'hypolimnion est régie par l'interaction de la température et du niveau de l'eau. La température et la teneur en PT sont en relation inverse avec la concentration initiale d'oxygène, mais l'effet du niveau de l'eau sur cette concentration initiale est aussi fonction de la température. Pour une température donnée, une augmentation du niveau de l'eau se traduit par une réduction de la concentration initiale d'oxygène, mais l'importance de cette réduction est fonction de la température. On a, pour illustrer ces points, reporté à la figure 61 les concentrations prévues par le modèle 5.10 en fonction des jours du calendrier, ceci pour plusieurs températures, niveaux d'eau et teneurs en PT. L'effet du niveau d'eau sur la concentration d'oxygène y apparaît comme important aux faibles températures de l'hypolimnion, mais peu important quand cette température se situe aux environs de 10°C. On y note aussi que, pour une valeur de PT de

0.01 et une période de stratification de 110 jours, il n'y a pas d'anoxie (concentration d'oxygène nulle) lorsque le niveau de l'eau est d'au moins 572 pieds; il y a probabilité d'anoxie lorsque le niveau d'eau tombe sous les 571 pieds.

Il s'avère enfin utile d'estimer la probabilité d'anoxie pour diverses périodes de stratification et teneurs en phosphore. Le temps nécessaire pour qu'il y ait anoxie est estimé en posant $y_{ij} = 0$ et en déterminant la valeur de $x_{ij}$ dans le modèle 5.10. Il est donc logique de supposer que la probabilité que la valeur de $y_{ij}$ soit inférieure ou égale à zéro corresponde à la probabilité d'anoxie, PA, qui est donnée par

$$\begin{aligned} PA &= P\{\hat{y}_{ij} \leq 0\} \\ &= P\{(\hat{y}_{ij} - \mu_{ij})/\sqrt{\underline{b}'C\underline{b}} \leq -\mu_{ij}/\sqrt{\underline{b}'C\underline{b}}\} \\ &= \int_{-\infty}^{-\frac{\mu_{ij}}{\sqrt{\underline{b}'C\underline{b}}}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \end{aligned} \qquad (5.15)$$

où C est la matrice des variances-covariances des coefficients du modèle 5.10, $\mu_{ij}$ est la valeur moyenne de $Y_{ij}$ et $\underline{b}$ est le vecteur colonne

$$\{1\ (x_{ij} - \bar{x})\ (x_{ij} - \bar{x})(w_j - \bar{w})\ (T_j - \bar{T})\ (T_j - \bar{T})(w_j - \bar{w})$$
$$(P_j - \bar{P})\ (x_{ij} - \bar{x})(P_j - \bar{P})\}$$

et $\underline{b}'$ est la transposée de $\underline{b}$. Il faut noter que, $\mu_{ij}$ étant fonction de la température, du niveau de l'eau et de la teneur en PT, PA est aussi fonction de ces variables. La valeur de $\mu_{ij}$ étant inconnue, on l'estime à partir de $\bar{y}_{ij}$.

On présente à la figure 62 les valeurs estimées de PA pour une période de stratification de 110 jours et deux teneurs en

phosphore (PT = 0.01 et PT = 0.02). Les valeurs du niveau de l'eau utilisées sont 570 pieds, 571 pieds et 572 pieds, et celles de la température de l'hypolimnion, 8°C, 9°C, 10°C, 11°C et 12°C. Pour PT = 0.01 et une température inférieure ou égale à 10°C, la probabilité d'anoxie est presque de 1 pour un niveau d'eau de 570 pieds, tandis que la valeur de PA est légèrement inférieure à 1 aux températures supérieures à 11°C. Aux niveaux d'eau de 571 pieds et 572 pieds, PA croît avec la température. Le passage de PT de 0.01 à 0.02 se traduit par une augmentation appréciable de la probabilité d'anoxie aux niveaux d'eau supérieurs à 571 pieds.

## CONCLUSION

On a élaboré, à l'aide des données recueillies de 1967 à 1979 par le Centre canadien des eaux intérieures, un modèle des concentrations d'oxygène dissous dans le bassin central du lac Erié. Le niveau de l'eau, la température de l'hypolimnion et la teneur en phosphore total sont les variables explicatives du modèle. On a trouvé que le taux d'appauvrissement est totalement indépendant de la température; il n'est fonction que du niveau de l'eau et de la teneur en PT. La concentration initiale d'oxygène de l'hypolimnion est cependant fonction de la température, de la teneur en PT et du niveau de l'eau. L'effet du niveau de l'eau est fonction de la température de l'hypolimnion, c'est-à-dire de l'interaction entre la température et le niveau de l'eau. Le modèle est utilisé pour dégager la tendance passée du taux d'appauvrissement après élimination des effets de la température et du niveau de l'eau. On conclut que l'accroissement du taux d'appauvrissement est lié à l'augmentation de la teneur en PT. Le modèle est, de plus, utilisé pour estimer la probabilité d'anoxie dans le bassin central en fonction des trois variables explicatives. On conclut aussi que la probabilité d'anoxie est toujours élevée et qu'elle croît avec la teneur en phosphore total. Il est donc possible de lutter contre l'anoxie du lac en réduisant la charge de PT.

Tableau 49. Estimations des paramètres du modèle 5.2

| Paramètre | Estimation | Paramètre | Estimation |
|---|---|---|---|
| $\beta_0$ | 1083.0931 | $\gamma_5$ | -0.089 |
| $\beta_1$ | -1063.618 | $\gamma_6$ | -0.084 |
| $\gamma_1$ | -0.118 | $\gamma_7$ | -0.066 |
| $\gamma_2$ | -0.119 | $\gamma_8$ | -0.106 |
| $\gamma_3$ | -0.106 | $\gamma_9$ | -0.087 |
| $\gamma_4$ | -0.113 | $\gamma_{10}$ | -0.095 |

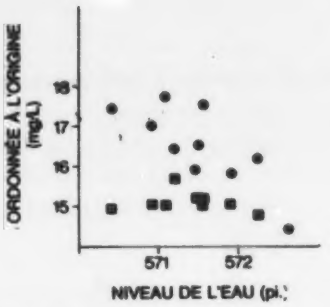



**Figure 55. Ordonnée à l'origine des droites de régression de la concentration d'oxygène dissous en fonction du niveau du lac Érié. Les carrés pour le bassin est et les cercles pour le bassin central.**

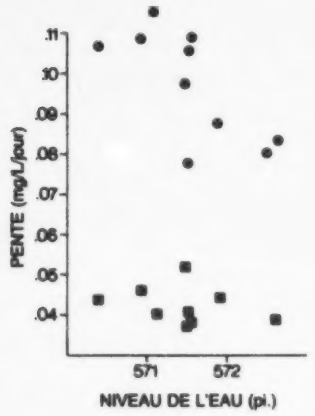



**Figure 56. Taux d'appauvrissement estimé en fonction du niveau du lac Érié. Les carrés pour le bassin est et les cercles pour le bassin central.**

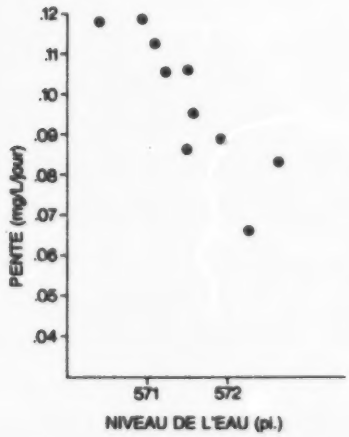



**Figure 57. Taux d'appauvrissement estimé selon le modèle 5.2 en fonction du niveau du lac Érié. Les carrés pour le bassin est et les cercles pour le bassin central.**

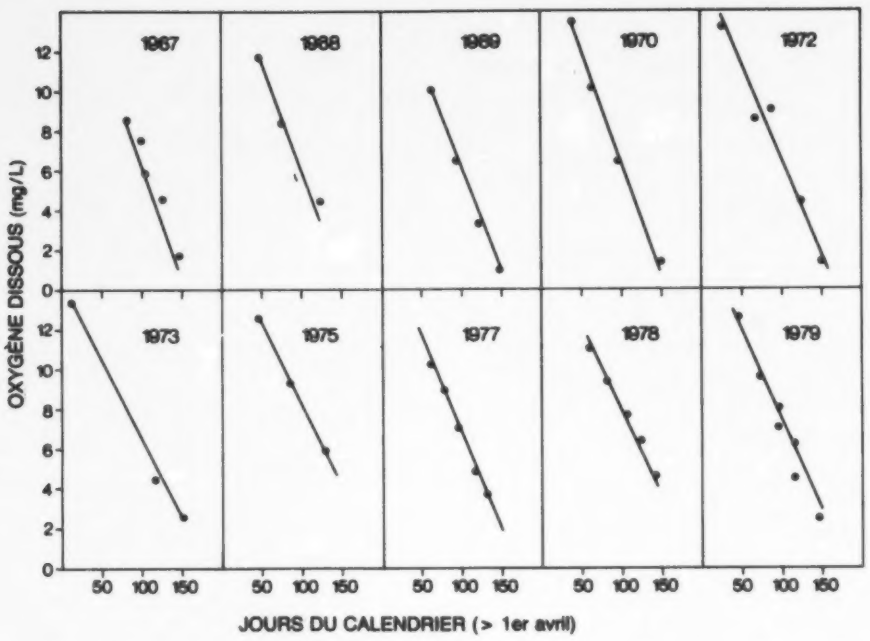



**Figure 58. Tracé des concentrations d'oxygène observées et de leurs valeurs estimées selon le modèle 5.10 en fonction du temps en jours du calendrier.**

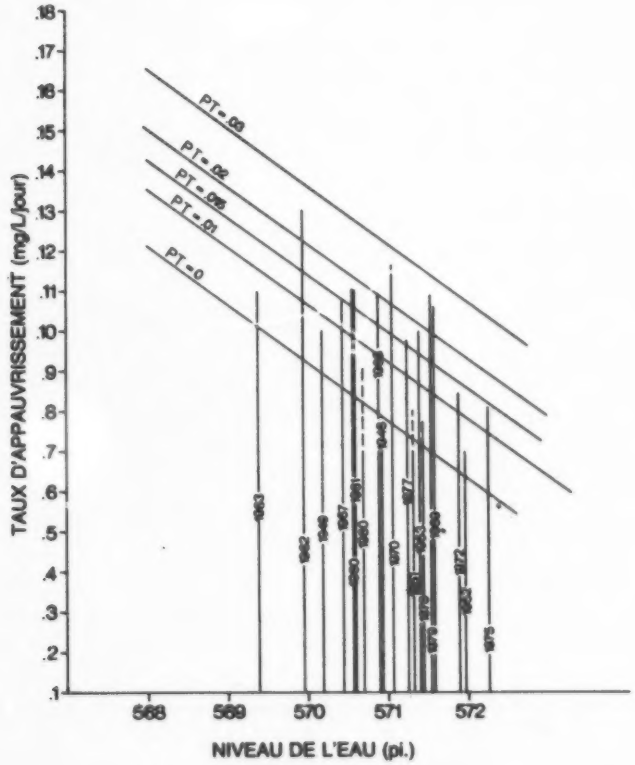



**Figure 59. Tracé du taux d'appauvrissement en fonction du niveau de l'eau pour diverses teneurs en PT.**

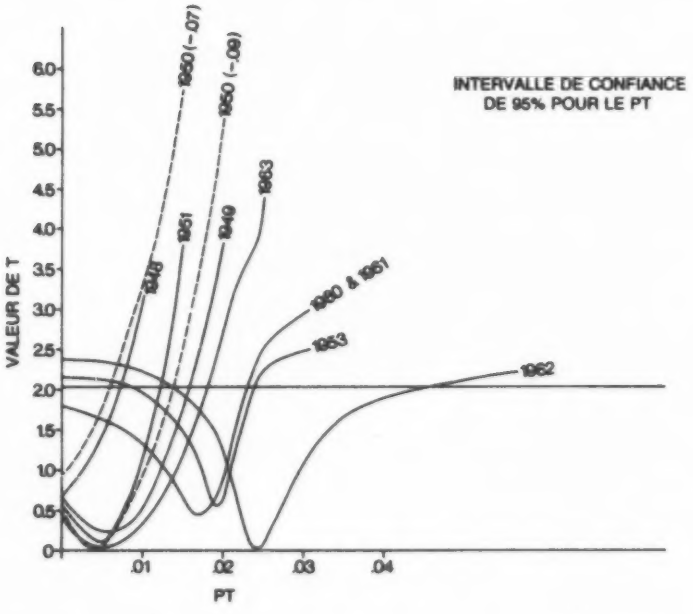



**Figure 60. Intervalle de confiance à 95 % pour les teneurs en PT, de 1948 à 1962.**

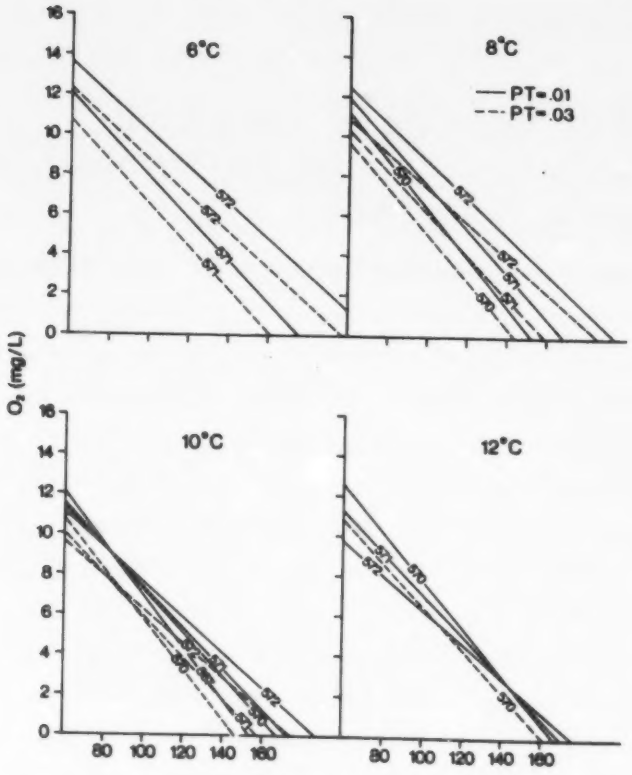



**Figure 61. Concentration de $O_2$ prévue par le modèle 5.10 en fonction des jours du calendrier pour plusieurs températures, niveaux de l'eau et teneurs en PT.**

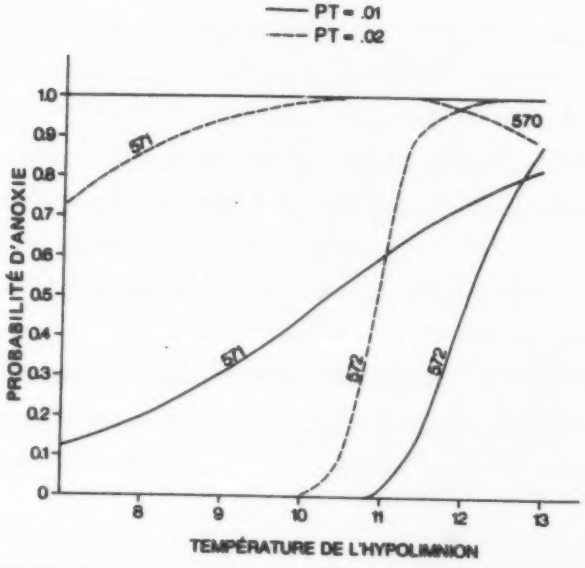



**Figure 62. Probabilité d'anoxie pour une période de stratification de 110 jours et diverses teneurs en PT et niveaux d'eau.**

# Distribution spatiale et temporelle des concentrations de coliformes totaux et fécaux, 1966-1970

*par S.R. Esterby et A.H. El-Shaarawi*

## INTRODUCTION

Ce chapitre porte sur la caractérisation des distributions spatiale et temporelle des concentrations de coliformes totaux et de coliformes fécaux dans le lac Erié à partir de données recueillies dans le cadre du programme de surveillance du ministère de l'Environnement (Centre canadien des eaux intérieures, 1966-1976). Les coliformes totaux et les coliformes fécaux sont couramment utilisés comme indicateurs bactériologiques de la qualité de l'eau; leur présence dans l'eau est considérée comme un indice de la présence d'organismes pathogènes. Ainsi, on fait mention, dans le document de Santé et Bien-être social Canada (1979) relatif à la teneur maximale acceptable pour l'eau potable traitée, de valeurs limites pour les concentrations de coliformes totaux et fécaux; de telles limites sont aussi mentionnées dans les lignes directrices relatives au traitement des eaux d'alimentation brutes. Les valeurs limites recommandées pour les eaux de baignade sont aussi basées sur des concentrations de coliformes fécaux (Environnement Canada, 1979). Même si les concentrations des échantillons prélevés à partir d'un bateau dans le cadre du programme de surveillance ne sont pas directement pertinentes à la plupart des lignes directrices ou des règlements qui visent particulièrement les usagers, elles doivent servir à évaluer la qualité bactériologique des eaux du large en fonction du nombre de coliformes.

Notre approche sera celle d'une analyse statistique des données visant à en expliquer la variation par des composantes systématique et aléatoire, cette dernière étant présentée sous la forme d'une distribution de probabilité. Cette analyse repose sur l'hypothèse voulant que les résultats des dénombrements, dans le temps ou l'espace, ont une distribution de Poisson (Feller, 1957). Il a été démontré que, dans des conditions bien contrôlées, les concentrations bactériennes d'échantillons dédoublés d'un milieu homogène obéissent à la loi de Poisson (Fisher et coll., 1922). Il est cependant généralement difficile d'appliquer cette loi simple comportant un seul paramètre à des concentrations bactériennes d'échantillons d'eau prélevés à des moments et endroits divers (Pipes et coll., 1977; El-Shaarawi et coll., 1981). L'analyse a donc été divisée en plusieurs étapes.

On a étudié la distribution des fréquences pour l'ensemble des données sur les coliformes totaux de chaque mission, et la distribution spatiale a été expliquée par une répartition par zones. Compte tenu des limites imposées par les données (tableau 50) et des incohérences dont nous traiterons plus loin, on a comparé les données d'une année à l'autre. On traite de la relation entre les concentrations des coliformes fécaux et celles des coliformes totaux, d'abord en établissant la distribution conditionnelle des fréquences des coliformes fécaux par rapport aux concentrations des coliformes totaux, puis en étudiant le rapport entre les coliformes fécaux et totaux. Comme les teneurs en coliformes totaux n'ont pas été relevées certaines années, l'étude de la constance des données au cours des années ne porte que sur les coliformes fécaux. Pour terminer, on traite aussi brièvement de la relation entre la turbidité et les concentrations de coliformes totaux.

Le présent chapitre comprend les sections suivantes: (a) méthodes statistiques, (b) coliformes totaux, (c) coliformes fécaux et (d) résumé et discussion.

## METHODES STATISTIQUES

On trouvera ci-après une description des principales méthodes statistiques qui seront utilisées dans les prochaines sections. L'analyse des données nécessitant, de par sa nature même, l'utilisation de techniques adaptées à chaque situation, quelques autres modes d'analyse seront présentés selon les besoins. Dans la présente section, on a choisi comme exemple les données de 1968 sur les coliformes totaux.

### Distributions de fréquences empiriques

En supposant qu'une variable aléatoire R a une distribution de probabilité P(R), on peut calculer la probabilité que R prenne certaines valeurs. Pour une série d'observations donnée, comme les concentrations de coliformes totaux d'une série d'échantillons d'eau, la distribution de fréquences empirique est analogue à la distribution de probabilités théorique. On obtient la distribution de fréquences empirique de la série de données sur les coliformes en comptant le nombre de fois que la concentration de coliformes prend chaque valeur distincte. Cette distribution peut être représentée graphiquement par un histogramme des fréquences dont la superficie de chaque bâtonnet est égale à la fréquence. On fait appel, dans le présent chapitre, à la distribution de fréquences empirique pour faciliter le choix d'une distribution de probabilités pertinente et pour comparer les distributions de diverses séries de données.

Soit $r_1$, $r_2 \ldots r_n$ le nombre d'organismes dans un volume normalisé pour n échantillons d'eau. La distribution de fréquences empirique est alors donnée par

$$f(r) = \frac{\text{Nombre de } r_i = r}{n} \qquad (6.1)$$

## Distributions de probabilités des concentrations bactériennes

L'application de la statistique à l'analyse des concentrations bactériennes est résumée par Pipes et coll. (1977) et El-Shaarawi et coll. (1981). Trois des lois les plus répandues sont la loi de Poisson, la loi binomiale négative et la loi lognormale. Soit R le nombre d'organismes par 100 mL d'eau. La probabilité que le nombre d'organismes soit r est alors donnée par

$$P(R=r) = e^{-\lambda} \lambda^r / r! \qquad (6.2)$$

si l'on suppose que R suit la loi de Poisson ou par

$$P(R=r) = \frac{(k+r-1)!}{r!\ (k-1)!} \quad \frac{p^r}{(1+p)^{k+r}} \qquad (6.3)$$

si R suit la loi binomiale négative. Si R suit la loi lognormale, S = ℓnR a alors une distribution normale et on peut utiliser les tables pour une distribution normale réduite.

Les lois binomiale négative et lognormale conviennent quand la variation est plus importante que celle observée pour une distribution de Poisson ajustée aux données et quand la distribution des fréquences est biaisée vers la droite. La moyenne et la variance de la loi de Poisson sont égales: $\mu = \sigma^2 = \lambda$. On peut s'attendre à ce que les concentrations bactériennes d'échantillons prélevés à des moments et dans des endroits différents présentent une plus grande variation. Fisher (1941) a montré que l'on obtient la distribution binomiale négative lorsqu'on suppose que la moyenne suit la loi gamma $\mu = \lambda$. La moyenne et la variance de la distribution binomiale négative ainsi obtenues sont respectivement: $\mu = pk$ et $\sigma^2 = p(1+p)k$. En faisant égaler les moyennes des distributions de Poisson et binomiale négative $\mu = \lambda = pk$, on obtient une

expression de la variance de la distribution binomiale négative plus un terme qui représente la variance en excès, soit $\sigma^2 = \lambda + \lambda^2/k$.

Si l'on désire une transformation des données qui stabilise la variance, une transformation logarithmique peut, comme cela a été montré (Brownlee, 1965), s'avérer pertinente si la variance est proportionnelle au carré de la moyenne. C'est pour obtenir une approximation de la normalité qu'on utilise la distribution lognormale pour les concentrations bactériennes. Cette distribution est souvent retenue parce que les propriétés de la distribution normale sont bien connues et les calculs sont faciles à effectuer. Les concentrations nulles compliquent cependant les choses, la distribution lognormale étant tout à fait inappropriée si le nombre de valeurs nulles est trop important. Il arrivera donc que l'une de ces deux distributions conviendra mieux aux données que l'autre.

Ajustement des distributions de probabilités et estimation des paramètres

Ces deux opérations ne peuvent être isolées, mais l'ordre de leur application à un ensemble de données est déterminé de façon logique. Il est possible, après avoir trouvé une distribution de probabilités qui convient aux données, de faire certains énoncés quant aux valeurs des paramètres. Le test d'ajustement consiste, tout d'abord, à poser une hypothèse voulant qu'une distribution particulière convienne aux données et, ensuite, à effectuer un test qui détermine à quel point l'hypothèse est fausse. Le choix d'une distribution de probabilités peut être basé sur les caractéristiques de la distribution des fréquences empirique ou sur l'expérience.

La statistique de précision de l'ajustement, $X^2$, (voir Snedecor et Cochran, 1967) peut servir à tester l'ajustement d'une distribution de probabilités à une série de données. On regroupe les

n observations en k classes en s'assurant qu'aucune des classes ainsi formées ne présente une fréquence espérée trop faible. On obtient

$$x^2 = \sum_{j=1}^{k} (f_j - e_j)^2 / e_j$$

où $f_j$ est la fréquence observée dans la classe j, $e_j$ est la fréquence espérée dans la classe j et où $X^2$ a une distribution du chi-carré approchée à (k-p-1) degrés de liberté, p étant le nombre de paramètres de la distribution. La règle d'application, donnée par Snedecor et Cochran, est que l'espérance peut être aussi faible que 1 pour deux classes à condition qu'elle ne soit pas inférieure à 5 pour la plupart des autres classes. C'est cette règle qui a été utilisée.

L'hypothèse testée étant composée, il est nécessaire d'estimer les paramètres de la distribution pour le calcul de l'espérance mathématique. On a utilisé des estimations par maximum de vraisemblance pour les paramètres p et k de la distribution binomiale négative car, pour les données sur les coliformes analysées ici, les estimations par la méthode des moments diffèrent considérablement de celles par la méthode du maximum de vraisemblance. On traitera plus loin de ce phénomène. On peut, pour tester l'ajustement de la distribution lognormale, calculer directement l'espérance à l'aide des tables de la distribution normale réduite. Si la moyenne et la variance de S = ln R sont, respectivement, représentées par $\mu$ et $\sigma^2$, $(S-\mu)/\sigma$ a alors une distribution N(0,1). En remplaçant $\mu$ et $\sigma^2$ par leurs estimations par maximum de vraisemblance, on obtient

$$\bar{S} = \left(\sum_{i-1}^{n} \ln R_i\right)/n \quad \text{et} \quad \hat{\sigma}_S^2 = \sum_{i=1}^{n} (\ln R_i - \bar{S})^2/(n-1),$$

$P(R \leq r) = P(\ln R \leq \ln r) = P(S \leq s)$ peut être obtenu.

Le test d'ajustement d'une distribution de Poisson à une série de données ne nécessite pas le regroupement de ces dernières. On peut ici utiliser l'indice de dispersion de Fisher (Fisher et coll., 1922)

$$D^2 = (n-1)\, s^2/\bar{r}$$

où $\bar{r}$ et s sont la moyenne et l'écart-type de n dénombrements $r_1$, $r_2 \ldots r_n$. La distribution de $D^2$ est bien approchée par une distribution du chi-carré à (n-1) degrés de liberté.

La moyenne de l'échantillon, $\bar{r}$, est l'estimation par la méthode des moments; elle est aussi l'estimation par maximum de vraisemblance de $\lambda$, le paramètre de la distribution de Poisson, donc, l'estimation de la moyenne et de la variance de la distribution. La fonction de vraisemblance relative ou les limites de confiance de $\lambda$ peuvent être utilisées comme mesure de l'incertitude dans l'estimation de $\lambda$. La fonction de vraisemblance relative de $\lambda$ est donnée par

$$R(\lambda; r_1, r_2 \ldots r_n) = L(\lambda; r_1 \ldots r_n) \,/\, L(\lambda; r_1 \ldots r_n)$$

où

$$L(\lambda; r_1, r_2 \ldots r_n) \propto \lambda^{\sum_{i=1}^{n} r_i} e^{-n\lambda}$$

et $\hat{\lambda}$ est l'estimation par maximum de vraisemblance de $\lambda$, c'est-à-dire la valeur de $\lambda$ qui maximise L (voir Sprott et Kalbfleisch, 1965 pour une étude sur l'utilisation de la fonction de vraisemblance). Les limites de confiance de $\lambda$ peuvent être obtenues à partir de la relation entre la distribution de Poisson et la distribution du $x^2$ (Brownlee, 1965). Les limites de confiance à $1-\alpha$ de $\lambda$ sont donc

$$\tfrac{1}{2}\chi^2_{1-\alpha/2}(2(r.+1)) \text{ et } \tfrac{1}{2}\chi^2_{\alpha/2}(2r.)$$

où

$$r. = \sum_{i=1}^{n} r_i \text{ et } \chi^2_p(\nu)$$

est la valeur de la variable aléatoire $\chi^2$ à $\nu$ degrés de liberté pour laquelle la fonction de distribution cumulative est égale à p. La valeur de $\chi^2_p(\nu)$ peut être obtenue directement des tables. Si r. est grand $\chi^2_p(\nu)$ peut être obtenu par approximation normale de la distribution $\chi^2$.

Les termes des deux premiers moments de la distribution binomiale négative sont $\mu = pk$ et $\sigma^2 = p(1+p)k$. En faisant égaler les moments d'échantillonnage et les moments théoriques, on obtient les estimations par la méthode des moments de p et de k.

$$\tilde{p} = \frac{s^2-\bar{r}}{\bar{r}} \quad \text{et} \quad \tilde{k} = \frac{\bar{r}^2}{s^2-\bar{r}}$$

où $\bar{r}$ et $s^2$ sont la moyenne de l'échantillon et la variance de $r_1$, $r_2 \cdots r_n$. Fisher (1941) a déterminé les conditions dans lesquelles les estimations par la méthode des moments sont efficaces. Elles sont très efficaces si l'une des conditions suivantes est respectée: (1) $\bar{r}$ est petit, p. ex. $\bar{r} = 0.1$ et $k>1$; (2) $\bar{r}=1$ et $k>3$; (3) $\bar{r} = 10$ et $k>9$; (4) $p<1/9$; ou (5) $k>18$. Les distributions binomiales négatives ajustées aux données sur les coliformes totaux présentaient des valeurs élevées de p et des valeurs faibles de k. Les données obtenues lors de la mission 102 de 1968 sont typiques à cet égard; l'efficacité des estimations par la méthode des moments n'était que de 0.63 environ. On devrait utiliser les estimations par maximum de vraisemblance lorsque la méthode des moments est peu efficace.

La fonction mixte de vraisemblance des paramètres p et k de la distribution binomiale négative est

$$L(p,k;r_1,r_2\ldots r_n) = \left[\prod_{i=1}^{n}\prod_{j=1}^{r_i}(k+r_i-j)\right] p^{n\bar{r}}/(1+p)^{n(k+\bar{r})}$$

Les estimations de maximum de vraisemblnce de p et k sont données par $\hat{p} = \bar{r}/\hat{k}$ où $\hat{k}$ est la solution de l'équation

$$\sum_{i=1}^{n}\sum_{j=1}^{r_i}\frac{1}{(k+r_i-j)} - n\,\ln(1+\bar{r}/k) = 0$$

qui doit être établie numériquement, ce qui est facile par ordinateur. La fonction mixte de vraisemblance relative

$$R(p,k;\ r_1,r_2\ldots r_n) = L(p,k;r_1,r_2\ldots r_n)/L(\hat{p},\hat{k};\ r_1\ldots r_n)$$

peut servir à déterminer des valeurs plausibles pour le couple de paramètres. Les paramètres peuvent être isolés à l'aide des fonctions de maximum de vraisemblance de p et k

$$L_m(p) = n\bar{r}\ln p - n(\hat{k}+\bar{r})\ln(1+p)$$

et

$$L_m(k) = \sum_{i=1}^{n}\sum_{j=1}^{r_i}\ln(k+r_i-j) - n(k+r)\ln(1+\hat{p})$$

Il faut noter, lors de l'interprétation de ces deux fonctions, que l'estimation par maximum de vraisemblance du deuxième paramètre apparait dans l'expression du premier. Ainsi, la fonction de maximum de vraisemblance de k permet d'obtenir les valeurs plausibles de k en supposant que p est égal à la valeur la plus plausible, c'est-à-dire $p = \hat{p}$.

Si R a une distribution lognormale à paramètres $\mu$ et $\sigma^2$, $S = \ln R$ a une distribution normale de moyenne $\mu$ et de

variance $\sigma^2$. Les estimations de $\mu$ et $\sigma^2$ par maximum de vraisemblance sont

$$\hat{\mu} = \frac{1}{n} \sum_{i=1} \ln r_i \quad \text{et} \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (\ln r_i - \hat{\mu})^2$$

La moyenne et la variance de R sont données par $\mu_r = e^{\mu+\frac{1}{2}\sigma^2}$ et et $\sigma_r^2 = e^{2\mu-\sigma^2} (e^{\sigma^2} - 1)$. Les estimations par la méthode des moments de $\mu$ et de $\sigma^2$ sont obtenues en faisant égaler les deux premiers moments d'échantillonnage de R.

$$\bar{r} = 1/n \sum_{i=1}^{n} r_i \quad \text{et} \quad s_r^2 = 1/(n-1) \sum_{i=1}^{n} (r_i - \bar{r})^2$$

à leurs moments théoriques correspondants, et en déterminant les valeurs de $\mu$ et de $\sigma^2$.

On obtient ainsi:

$$\tilde{\mu} = 2 \log \bar{r} - \tfrac{1}{2} \log (s_r^2 + \bar{r}^2)$$

$$\tilde{\sigma}^2 = \log (s_r^2 + \bar{r}^2) - 2 \log \bar{r}$$

Aitchison et Brown (1957) ont cependant montré que l'efficacité des estimateurs de la méthode des moments diminuait à mesure que $\sigma^2$ augmentait et que, même pour une valeur de $\sigma^2$ aussi petite que 0.5, l'efficacité de $\tilde{\mu}$ n'était que de 79 % et celle de $\tilde{\sigma}^2$ que de 31 %. Mais comme l'indiquent les données sur les coliformes totaux plus loin dans ce chapitre, la valeur de $\sigma^2$ est probablement beaucoup plus grande que 0.5, la plus petite valeur observée de l'estimation efficace de maximum de vraisemblance ayant été $\tilde{\sigma}^2 = 0.98$.

Les estimations de $\mu$ et de $\sigma^2$ conviennent quand on ne désire effectuer qu'un test d'ajustement. Il faut cependant effectuer des calculs supplémentaires pour obtenir des estimations de la moyenne et de la variance de R. Les estimations de la méthode des moments sont

$$\tilde{\mu}_r = \bar{r} \quad \text{et} \quad \tilde{\sigma}_r^2 = \frac{n-1}{n} s_r^2$$

mais, encore ici, l'efficacité des estimations est faible. Des approximations des estimations par maximum de vraisemblance pour un échantillon important ont été données par Aitchison et Brown (1957)

$$\hat{\mu}_r = e^{\bar{s}} \Psi_n (\tfrac{1}{2}s_s^2)$$

$$\hat{\sigma}_r^2 = e^{2\bar{s}} \{\Psi_n(2s_s^2) - \Psi_n (\frac{n-2}{n-1} s_s^2)\}$$

et

$$\Psi_n(t) \simeq e^t \{1 - t \frac{(t+1)}{n} + t^2 (\frac{3t^2+22t+21}{6n^2})\}$$

où $s_s^2 = \frac{n}{n-1} \hat{\sigma}^2$. Il arrive parfois qu'une ligne directrice sur la qualité de l'eau précise une moyenne géométrique. Ainsi, il est recommandé (Environnement Canada, 1979) que la teneur en coliformes fécaux des eaux de baignade ne dépasse pas 200 bactéries par 100 mL dans un minimum de cinq échantillons prélevés au cours d'une période de 30 jours. On présume ici que la concentration de coliformes fécaux, R, a une distribution lognormale. La médiane d'une distribution lognormale à paramètres $\mu$ et $\sigma^2$ est $e^{\mu}$. L'estimation par maximum de vraisemblance de la moyenne d'un échantillon $r_1, r_2 \ldots r_n$ est $e^{\hat{\mu}}$ mais $e^{\hat{\mu}} = e^{\bar{s}}$ = moyenne géométrique, car

$$\bar{s} = 1/n \sum_{i=1}^{n} \ln r_i$$

La distribution lognormale est définie pour $r > 0$; les concentrations nulles posent donc un problème. Le fait que $R = r$ et que $P\ (r-\delta<R<r+\delta)$ offrent deux solutions possibles. Si l'on définit une variable $y = \ln(R+1)$ dont la distribution de probabilités est donnée par

$$P(Y < 0) = 0$$
$$P(Y = 0) = P(0 < Y < 1) = \int_0^1 f(y)dy$$
$$P(Y \leq y) = \int_0^y f(y)dy \text{ pour } (y > 0)$$

où $f(y)$ est la fonction de densité de probabilité d'une distribution normale de moyenne $\mu$ et de variance $\sigma^2$, le problème de l'estimation est donc réduit à la solution d'une distribution normale tronquée à l'origine. La deuxième solution consiste à supposer que seules les valeurs non nulles ont une distribution lognormale avec des proportions de valeurs nulles $\delta$ et de valeurs non nulles $n-\delta$. On obtient ainsi une distribution $\Delta$ à trois paramètres qui sont $\delta$ = la proportion de valeurs nulles, et $\mu$ et $\delta^2$, les paramètres de la distribution lognormale. Aitchison et Brown ont établi les estimateurs de maximum de vraisemblance des paramètres ainsi que la moyenne et la variance. Une solution pratique consiste à ajouter 1 à toutes les concentrations et à procéder comme si R présentait une distribution lognormale. Si le nombre de valeurs nulles est faible et si les observations non nulles sont élevées, la distortion des estimations est alors faible.

Méthodes par regroupement en zones

Nous avons jusqu'à maintenant supposé qu'une seule distribution de probabilités permettait de décrire adéquatement les concentrations bactériennes de tout un lac. On peut mieux résumer

les concentrations bactériennes d'un lac en divisant celui-ci en zones homogènes, c'est-à-dire des zones où les concentrations bactériennes peuvent être décrites par la même distribution de probabilités, celle-ci différant cependant d'une zone à l'autre. Comme les concentrations bactériennes d'un échantillon homogène obéissent à la loi de Poisson, il est logique d'utiliser une méthode permettant de diviser l'ensemble des données en sous-ensembles; les sous-ensembles ont des distributions de Poisson de moyennes différentes. Une telle méthode, basée sur l'indice de dispersion $D^2$, est présentée dans El-Shaarawi et coll. (1981). On pourra trouver une version améliorée de cette méthode dans Esterby et El-Shaarawi (1981) et, dans El-Shaarawi (1982), une généralisation de la méthode à d'autres distributions au sein des sous-ensembles et un autre critère d'arrêt.

On a, dans le présent chapitre, utilisé la méthode de regroupement en grappes décrite par Esterby et El-Shaarawi (1981). Soit k le nombre de groupes et $\lambda_1, \lambda_2 \ldots \lambda_k$ les moyennes des distributions de Poisson de chacun des k groupes. Une observation r est ensuite assignée au groupe g si $L(r) = P(r;\lambda_g)/P(r;\lambda_j) > 1$ pour $j \neq g$; $P(r;\lambda)$ est la fonction de probabilités de Poisson. On assigne ainsi une observation au sous-ensemble ou au groupe pour lequel la probabilité que la variable de Poisson prenne cette valeur observée est la plus grande. Il s'agit donc d'une méthode de regroupement par probabilités suivant deux possibilités: (1) répartition optimale dans un nombre de groupes fixes et (2) choix du nombre de groupes. L'application de cette méthode comporte plusieurs étapes. Le nombre de groupes, k, est fixé à sa valeur minimale. On assigne des valeurs initiales aux paramètres $\lambda_1 \lambda_2 \ldots \lambda_k$ où $\lambda_i$ est la moyenne de la distribution de Poisson du $i^{ième}$ groupe et $\lambda_1 < \lambda_{i+1}$. Les valeurs initiales $s_1, s_2 \ldots s_k$ sont obtenues en divisant la

gamme des valeurs observées en k intervalles égaux et en posant $s_i$ le point milieu du $i^{ième}$ intervalle. On détermine ensuite $c_1$, $c_2 \ldots c_{k-1}$ où $c_i$ est la valeur de r qui satisfait l'équation $L(r) = P(r;\lambda_{i+1})/P(r;\lambda_i)=1$ avec $P(r;\lambda) = \lambda^r e^{-\lambda}/r!$. Toutes les observations $r_1$, $r_2 \ldots r_n$ sont assignées aux groupes appropriés par comparaison avec $c_1$, $c_2 \ldots c_k-1$. Si $r_j$, $r_j > c_i$, $r_j$ est assigné au i+1 groupe, la solution de cette équation peut être réduite en

$$c_i = (\lambda_{i+1} - \lambda_i)/(\log \lambda_{i+1} - \log \lambda_i)$$

Une fois toutes les observations assignées aux groupes, les paramètres $\lambda_1$, $\lambda_2 \ldots \lambda_k$ sont estimés à partir de la moyenne des échantillons du groupe. S'il s'agit de la première itération, on recalcule les valeurs $c_1$, $c_2 \ldots c_{k-1}$ et on assigne de nouveau les échantillons aux groupes. On continue d'assigner aux groupes et de recalculer les valeurs de $\lambda$ jusqu'à ce que la composition des groupes reste la même lors de deux itérations successives.

Le nombre minimal de groupes devrait normalement être de 2. On calcule, pour chaque valeur de k, les indices de dispersion $D_1$, $D_2 \ldots D_k$ et les niveaux de signification correspondants $\alpha_1$, $\alpha_2 \ldots \alpha_k$. Une valeur de $D^2$ significative, indiquant une non homogénéité dans au moins un des groupes, signifie que k doit être augmenté et que la répartition des échantillons parmi les groupes doit être répétée. Ainsi, pour tout k, le critère d'arrêt est min $(\alpha_1, \alpha_2 \ldots \alpha_k)$ inférieur à un niveau de signification arbitraire.

La généralisation de la méthode décrite par El-Shearawi (1982) permet de donner à $P(r_1;\lambda)$ d'autres formes que celle de la distribution de Poisson. Le critère d'arrêt est différent et basé sur le rapport de vraisemblance

$$\Lambda = \{ \prod_{i=1}^{k} \prod_{j=1}^{n_i} P(r_j;\lambda_i^{(k)}) \} / \{ \prod_{i=1}^{k+1} \prod_{j=1}^{n_i} P(r_j;\lambda_i^{(k+1)}) \}$$

où $\lambda_i^{(k)}$ est la valeur de $\lambda$ du $i^{ième}$ groupe avec k groupes et où $\lambda_i^{(k+1)}$ est la valeur de $\lambda$ du $i^{ième}$ groupe avec k+1 groupes. Il s'agit donc du rapport des probabilités des observations selon les deux hypothèses, à savoir avec k et k+1 groupes. On peut, pour tester la nécessité de k+1 plutôt que k groupes, utiliser le terme -2 log $\Lambda$ qui, pour n élevé, a une distribution $x^2$ à un degré de liberté.

La méthode de regroupement utilisée jusqu'à maintenant ne tient pas compte de la position des échantillons dans l'espace. Pour définir des zones, on indique le groupe auquel une station appartient sur une carte du lac. Il semblait plus logique, lorsqu'on s'est intéressé aux méthodes de répartition en zones, d'inclure la position géographique des stations à la méthode même du regroupement. Un critère de zonage basé sur la distance entre les stations a eu pour effet de définir trop de groupes. Il est apparu en comparant les résultats avec ceux de la méthode de El-Shaarawi et coll. (1981), que ce phénomène était dû à la nature discontinue des zones le long de la rive. Cette discontinuité interdisait aussi l'utilisation de méthodes de projection de surfaces ou de courbes de niveau qui sont applicables à des variables aléatoires continues (El-Shaarawi et Esterby, 1981). Une zone est donc formée de régions du lac, non nécessairement continues, qui ont été regroupées. Elles présentent des concentrations bactériennes que l'on peut considérer comme ayant la même distribution de probabilités; ainsi, la même moyenne peut être attribuée à toutes les stations d'une même zone.

Un exemple: l'analyse des données de 1968 sur les coliformes totaux

Des données sur les coliformes totaux ont été recueillies au cours de quatre missions effectuées en 1968 sur le lac Erié (tableau 50). Le nombre de stations visitées au cours de la mission 102 a été

de 63, mais seuls les échantillons prélevés à la profondeur de 1 m ont fait l'objet d'une analyse bactériologique systématique. On indique au tableau ci-après la profondeur et le nombre de stations pour lesquels les concentrations de coliformes totaux ont été notées.

| Profondeur | 1 | 7 | 8.5 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Nombre | 62 | 32 | 1 | 2 | 5 | 6 | 5 | 4 | 1 | 3 | 3 |

| Profondeur | 17 | 18 | 19 | 20 | 21 | 22 | 27 | 28 |
|---|---|---|---|---|---|---|---|---|
| Nombre | 1 | 3 | 3 | 2 | 2 | 2 | 1 | 1 |

La même situation s'étant répétée lors des autres missions, on a décidé de ne pas étudier la relation entre la concentration et la profondeur. Seules les concentrations à 1 m ont été analysées, ceci pour éviter de confondre les variations spatiales verticales et horizontales.

Une distribution binomiale négative a semblé bien décrire les données de chacune des quatre missions; ce qui n'était pas le cas de la distribution de Poisson. La proportion élevée de valeurs nulles, à savoir 29/62, 13/62, 10/28 et 26/52, pour les missions 102, 108, 109 et 112 respectivement, interdisait l'utilisation de la distribution lognormale. On présente à la figure 63 les fréquences observées et ajustées en fonction de la concentration allant de 0 à 18 coliformes par 100 mL. La distribution de Poisson et la distribution binomiale négative sont discrètes et les fréquences devraient être représentées par des bâtonnets, mais on les a représentées par des courbes pour les distinguer facilement des fréquences observées chaque année. Les distributions de fréquences avec celles de toutes les autres années paraissent à la figure 67. Les distributions sont fortement biaisées; la valeur modale r est égale à 0.

Les tests d'ajustement de la distribution binomiale négative aux données des quatre missions de 1968 ont été effectués par les méthodes des moments et du maximum de vraisemblance permettant d'estimer p et k dans le calcul de la statistique de précision de l'ajustement. Les fréquences observées et prévues, regroupées de façon à respecter le critère qui admet deux fréquences prévues aussi faibles que 1 à condition que les autres soient d'au moins 5, sont présentées à la partie supérieure du tableau 52. La statistique de précision de l'ajustement, le niveau de signification et les estimations de p et k sont présentés au bas du même tableau. Dans le cas des quatre missions, l'estimation de p par la méthode des moments est supérieure à l'estimation par maximum de vraisemblance, et l'on note l'inverse pour k. L'ajustement est meilleur, pour les données de chaque croisière, avec les estimations de maximum de vraisemblance. Le cas extrême est celui de la mission 108 où le niveau de signification de 0.08 révèle un ajustement médiocre des estimations de la méthode des moments comparativement à un niveau de signification de 0.74 pour les estimations par maximum de vraisemblance. Une surestimation du nombre de valeurs nulles est la cause d'un écart systématique entre les fréquences observées et prévues quand on utilise les estimations de la méthode des moments.

Les fonctions de vraisemblance relative mixtes pour p et k sont présentées à la figure 64 sous forme d'isolignes de vraisemblance relative où $R(p,k) = 0.01, 0.05, 0.50, 0.90$ et $0.99$. Les fonctions de maximum de vraisemblance relative de p, $R_m(p)$, et de k, $R_m(k)$, sont présentées aux figures 65a et 65b. Les isolignes de vraisemblance mixtes sont de forme triangulaire, la gamme des valeurs plausibles de p diminuant à mesure que celle de k s'accroît. Les énoncés relatifs à la plausibilité de paires de valeurs p et k devraient être basés sur les valeurs de $R(p,k)$ et non sur $R_m(p)$ et $R_m(k)$. Ainsi, lors de la mission 102, les valeurs de k où $k > 0.32$ ou $k < 0.12$ sont fortement improbables quand $p = \hat{p} = 32.58$ car $R(\hat{p},k) < 0.01$ pour $k > 0.32$ ou $k < 0.12$ quand $p = 97.31$,

c'est-à-dire $R(p,k) < 0.10$ pour $k > 0.19$ et $k < 0.11$. Des résultats semblables peuvent être obtenus pour p. Les estimations $\tilde{p}$ et $\tilde{k}$ obtenues par la méthode des moments ont été indiquées à la figure 64 pour la mission 102; on peut noter que $R(\tilde{p},\tilde{k})$ se situe près de l'isoligne 0.50.

Voyons maintenant le cas des distributions binomiales négatives dont les paramètres égaux aux estimations de maximum de vraisemblance obtenues pour ces quatre missions. On a signalé à la section précédente que la variance de la distribution binomiale négative $\sigma^2 = \lambda + \lambda^2/k$ où $\lambda = pk$ est la moyenne; le terme $\lambda^2/k$ a été nommé la variance en excès de la loi de Poisson dont la variance est égale à la moyenne. La substitution par $\tilde{p}$ et $\tilde{k}$ dans l'expression de la variance permet donc d'obtenir le terme $\lambda^2/k$ exprimé comme un pourcentage de la variance totale, soit 97, 98, 97 et 98 pour cent pour les missions 102, 108, 109 et 112 respectivement. On a fait appel aux méthodes de regroupement en zones pour tenir compte de cette variation en excès.

Les méthodes de regroupement décrites à la section précédente ne définissent des zones que pour une seule mission; il est donc nécessaire de trouver un moyen de comparer les zones de missions différentes. On peut décider que les zones de toutes les missions seront celles qui auront été déterminées pour une mission de référence donnée. La difficulté de cette approche réside en ce que le nombre et l'emplacement des stations varient d'une mission à l'autre. Une autre possibilité consiste à fixer le nombre de zones; les conséquences de cette approche sont étudiées ci-après à l'aide des données des missions de 1968.

On présente au tableau 53 les groupes déterminés pour la mission 102 à l'aide de la méthode de regroupement par probabilités faisant appel à l'indice de dispersion et au critère d'arrêt $-2 \log \Lambda$ ainsi que les groupes déterminés par l'indice de

dispersion. La partie supérieure du tableau donne pour les nombres de groupes k = 2, 3, 4, 5, 6 et 7: la moyenne du groupe, le nombre de stations dans le groupe et les valeurs de l'indice de dispersion et du niveau de signification telles qu'obtenues par regroupement en grappes des probabilités. La valeur de -2 log $\Lambda$ est aussi présentée pour k = 2, 3, 4, 5 et 6. Les derniers groupes, obtenus à l'aide de l'indice de dispersion, sont présentés à la dernière rangée du tableau. Dans le regroupement selon l'indice de dispersion, il suffit de quatre groupes pour assurer qu'aucune des valeurs de $D^2$ ne dépasse de 1 % la valeur critique de $x^2$. Le regroupement par probabilités nécessite, quant à lui, sept groupes si l'on fait appel au critère d'arrêt $D^2$ et six groupes si on utilise -2 log $\Lambda$; ici aussi la valeur critique correspond au niveau de signification de 1 %. La comparaison des groupes obtenus avec les deux méthodes quand k=4 montre que l'écart est dû à l'affectation des stations dans les deux premiers groupes; les moyennes sont de 0.44 et 7.14 avec 43 et 14 stations dans les deux premiers groupes, si l'on utilise le regroupement par probabilités et de 0.57 et 7.83 avec 45 et 12 stations quand on utilise l'indice de dispersion. Il faut aussi noter que le nombre de groupes passe de trois à six quand on utilise le regroupement par probabilités; aucune réaffectation ne touche les groupes 1 et 2. Les groupes 1 et 2 ne se fractionnent que si k = 7.

Pour évaluer les conséquences de ces écarts, on a reporté à la figure 66 les fonctions de probabilité assimilées à des distributions de Poisson à paramètre égal aux moyennes des groupes. La séparation des groupes est une caractéristique que devrait présenter toute méthode de regroupement; mais elle n'a pas été explicitement incorporée dans les méthodes utilisées. Dans le présent cas, cela peut signifier que les distributions de probabilités des groupes n'ont pas de gammes de valeurs chevauchantes dont la probabilité est non négligeable. Le chevauchement peut être étudié dans cette figure. On présente à la figure 66a les fonctions de probabilité basées sur les quatre groupes obtenus par regroupement

par probabilités; la distribution des fréquences observées est présentée en dessous. Ici aussi, pour faciliter la présentation et la séparation visuelle des fonctions de probabilité chevauchantes, les points de probabilité non nulle, c'est-à-dire les points (r, $P(r; \lambda)$) pour r = 0,1,2..., ont été reliés par des traits, mais il faut se rappeler que $P(x;\lambda) = 0$ pour $r < x < r + 1$, où r est un nombre entier. On présente à la figure 66b les fonctions de probabilité basées sur les deux premiers groupes obtenus à l'aide de l'indice de dispersion. A noter que les groupes 3 et 4 sont les mêmes quelle que soit la méthode. Le chevauchement étant essentiellement le même aux figures 66a et 66b, les mêmes conclusions peuvent donc être tirées pour la série de données. On note cependant, à la figure 66c, que le fractionnement des groupes 1 et 2 à n=43 et n=14 se traduit par trois distributions qui se chevauchent sensiblement. Il est donc préférable d'utiliser le critère d'arrêt $-2\lambda \log \Lambda$ plutôt que le critère $D^2$ pour cette série de données. La dernière question à résoudre consiste à savoir si le fractionnement des groupes 3 et 4, basés sur k=4, en 4 groupes avec n=1, 1, 1, et 2 quand k=7 justifie des calculs supplémentaires. On a décidé que cela n'était pas justifié étant donné la très grande dispersion des stations à ces concentrations élevées. Les résultats des trois autres missions de 1968 ont été étudiés de la même façon que ceux de la mission 102. Ici encore la séparation en un nombre plus élevé de groupes avait tendance à déplacer les stations des groupes 3 et 4 pour k=4. Il y avait cependant, pour chacune de ces trois missions, plus de chevauchement des fonctions de probabilité avec $D^2$ qu'avec la méthode de regroupement. L'approche définitive choisie pour la division en zones de toutes les données sur les coliformes totaux a donc consisté à fixer le nombre de zones à k=4 et à utiliser la procédure de regroupement par probabilités.

Une comparaison qualitative des zones des missions de 1968 est présentée dans El-Shaarawi et coll. (1981). Bien que l'on ait utilisé l'indice de dispersion par le regroupement et que le nombre

de zones ait été de 4, 5, 3 et 4 respectivement pour les missions 102, 108, 109 et 112, les observations sont très semblables à celles faites suite au regroupement en zones présenté au tableau 58 et à la figure 69. Ce dernier regroupement est basé sur 4 zones et consiste en un regroupement par probabilités. Ces zones de concentrations plus élevées forment des régions discontinues le long de la rive et seules celles de la mission 108, effectuée du 29 juillet au 3 août, s'étendent sur tout le lac.

Inférence relative au rapport des concentrations bactériennes

Comme les concentrations de coliformes totaux regroupent des bactéries d'origine fécale et non fécale et que la concentration de coliformes totaux sert à indiquer la présence d'organismes pathogènes, il est important de connaître le rapport entre les concentrations de ces deux types de coliformes. Le choix d'un organisme indicateur, comme les coliformes totaux, a toujours nécessité que l'indicateur soit plus facile à déceler que le pathogène, de par son nombre plus élevé. L'évaluation des risques pourrait être fortement améliorée si l'on pouvait déterminer un rapport stable entre les organismes pathogènes et les indicateurs; l'inférence portant sur le rapport de deux concentrations bactériennes est donc importante (El-Shaarawi et Pipes, 1982). El-Shaarawi et Pipes ont comparé le rapport des moyennes de deux distributions de Poisson indépendantes et celui des médianes de deux distributions lognormales.

L'inférence relative au rapport entre les concentrations de coliformes fécaux et de coliformes totaux peut s'imposer lorsque les données sont recueillies dans des circonstances différentes: (1) les concentrations de coliformes fécaux et totaux sont déterminées à partir des mêmes échantillons d'eau, mais sous forme de mesures indépendantes; (2) comme en (1), mais sous forme de mesures dépendantes, l'identification des coliformes fécaux étant faite à

partir de la plaque des coliformes totaux et (3) on dispose de séries individuelles de données sur les coliformes fécaux et totaux présents dans la même eau pour l'estimation du rapport. Ce dernier est nettement moins favorable que les deux premiers.

Il est possible d'appliquer les résultats présentés par El-Shaarawi et Pipes (1982) en supposant que les concentrations de coliformes totaux et fécaux sont dépendantes. Supposons que F, la concentration de coliformes fécaux et O, la concentration de coliformes non fécaux, présentent des distributions de Poisson indépendantes de moyennes respectives $\lambda_f$ et $\lambda_o$. L'inférence relative au rapport $\rho=\lambda_f/\lambda_o$ peut alors être basée sur la distribution conditionnelle de F où T=F+O, et T étant la concentration de coliformes totaux. On obtient, à partir de $f_1, f_2 \ldots f_n$ et $t_1, t_2 \ldots t_n$, où $o_i=t_i-f_i$, l'estimation par maximum de vraisemblance de $\rho$, en utilisant la fonction de vraisemblance conditionnelle:

$$\hat{p} = \sum_{i=1}^{n} f_i \; / \; \sum_{i=1}^{n} o_i$$

Il apparait clairement que le rapport $\rho = \lambda_f/\lambda_o$ ne peut être estimé par application directe des résultats de El-Shaarawi et Pipes si T et F sont indépendants, car il y a possibilité de valeurs négatives de O. Si T et F sont indépendants et que T suit une loi de Poisson de moyenne $\lambda_t$, le rapport $P_1 = \lambda_f/\lambda_t$ peut être estimé par la méthode ci-dessus. On obtient

$$\hat{P}_1 = \sum_{i=1}^{n} f_i \; / \; \sum_{i=1}^{n} t_i$$

Cependant, si les concentrations de coliformes totaux et fécaux sont déterminées à partir du même échantillon, on se retrouve

dans la situation de couples appariés. Si T et F sont indépendants et présentent une distribution de Poisson, il s'avère plus pertinent de supposer que la moyenne des $T_i$ est une fonction d'un paramètre spécifique aux couples qui n'est liée qu'à la concentration de coliformes totaux; on procède de même pour F. Si la moyenne est une fonction multiplicative, on obtient

$$\mu_{t_i} = \beta_i \lambda_t \quad \text{et} \quad \mu f_i = \beta_i \lambda_f$$

où $\mu_{ti}$ est la moyenne des $T_i$, $\mu f_i$ la moyenne des $F_i$ et $\beta_i$ le paramètre spécifique au couple. On obtient pour $\rho = \lambda_f/\lambda_t$

$$P(t_i) = \theta_i^{t_i} \frac{e^{-\theta_i}}{t_i!} \qquad P(f_i) = (\rho\theta_i)^{f_i} \frac{e^{-\rho\theta_i}}{f_i!}$$

où $\theta_i = \beta_i\lambda_t$ et, conséquemment, $\mu_{f_i} = \beta_i\lambda_f = \beta_i\lambda_t\rho = \theta_i\rho$. Or il s'agit d'un cas particulier de la distribution d'une série puissance étudié par Kalbfleisch et Sprott (1973) qui donne la vraisemblance conditionnelle

$$\prod_{i=1}^{n} \binom{t_i + f_i}{f_i} \rho^{\sum_{i=1}^{n} f_i} / (1+\rho)^{\sum_{i=1}^{n} (t_i + f_i)}$$

et $\hat{\rho} = \sum_{i=1}^{n} f_i / \sum_{i=1}^{n} t_i$. Ce résultat est le même que celui que l'on a obtenu précédemment à l'aide des résultats de El-Shaarawi et Pipes (1981) en supposant que T et F sont indépendants. Si $\mu_{t_i} = \beta_i + \lambda_t$ et $\mu_{f_i} = \beta_i + \lambda_f$ l'estimation de $\rho = \lambda_f/\lambda_t$ s'avère alors plus difficile.

## ANALYSE DES DONNEES SUR LES COLIFORMES TOTAUX

Des données sur les coliformes totaux ont été recueillies lors des 16 missions effectuées sur le lac Erié entre août 1966 et août 1970; elles sont cependant mal réparties entre les années, et le nombre et les emplacements des stations sont très variables d'une mission à l'autre. L'analyse des données effectuée dans la présente section sera faite dans l'ordre de la présentation des méthodes déjà décrit.

On pourra trouver à la figure 67 la distribution des fréquences observées pour chacune des 16 missions; les missions sont disposées par ordre chronologique de haut en bas et de gauche à droite jusqu'à la page suivante. Les flèches numérotées indiquent que la valeur observée déborde de l'histogramme. Les distributions de fréquences sont présentées en deux échelles dans des histogrammes adjacents. Pour les missions effectuées en 1967, 1968 et 1969, le côté gauche de l'histogramme correspond à la gamme des observations, ce qui explique que la plupart de ces histogrammes présentent un pic situé à gauche et des valeurs beaucoup plus faibles clairsemées le long de l'axe. Ces tracés ont été inclus afin de souligner l'extrême biais des données et de permettre la comparaison des données moins biaisées de 1968 avec celles de 1967 et 1969. Le côté droit des histogrammes, dont la gamme des concentrations va de 0 à 14 ou de 0 à 16, indique que la valeur modale est 0 et que, même pour cette gamme, la distribution est très biaisée. Les deux histogrammes de droite en lignes hachurées présentent la fréquence de chaque intervalle de 5. L'histogramme de gauche des missions de 1966 et 1970 est le même que celui décrit plus haut, mais l'histogramme de droite est celui de la distribution des fréquences du logarithme naturel de la concentration. Comme cela a été mentionné auparavant, la transformation logarithmique a été utilisée afin de satisfaire l'hypothèse de normalité. La distribution des fréquences ainsi obtenue pour la mission de 1970 est beaucoup plus symétrique que

celle des données non transformées. On note cependant que, pour la mission 111 de 1966, la distribution des fréquences est encore biaisée et semble bimodale. Les histogrammes sont basés sur un intervalle de regroupement de 0.5 et de 1.0 $\log_e R$. Le pic important noté dans l'intervalle [0, 0.5] correspond à la concentration zéro. On a, pour les données de 1966, ajouté 1 à toutes les concentrations avant d'effectuer la conversion logarithmique.

On peut voir, à partir des distributions de fréquences observées, que les données se divisent en deux séries de types fondamentalement différents: une regroupant les données de 1967, 1968 et 1969 et l'autre celles de 1966 et 1970. La distribution des fréquences des missions de 1967 et 1968 présente une allure semblable. La plupart des concentrations se situent dans la gamme de 0 à 100 coliformes par 100 mL; il y a un grand nombre de valeurs nulles et un biais marqué, même dans la gamme de 0 à 16, et très peu de concentrations sont supérieures à 100. Cette caractéristique apparait même si le plan d'échantillonnage est variable, comme cela est indiqué par les emplacements des stations présentés à la figure 68. Il faut cependant noter que les distributions des fréquences des missions 107, en 1967 et 102, 108, 109 et 112, en 1968, sont basées sur une gamme de 0 à 100 pour les histogrammes de gauche, tandis qu'une gamme de 0 à 1000 a été utilisée pour toutes les autres missions de ces années.

Les données de 1969 ont été enregistrées d'une manière totalement différente. Il y a eu beaucoup de concentrations nulles en 1966, 1967 et 1968. Au contraire, des valeurs de <1, <2 et <3 ont été enregistrées. Ces valeurs indéterminées ont été modifiées en zéros afin de standardiser les données de 1969. Cette approche apparait très logique pour les <1. Quant aux <2 et <3, ils ont été remplacés par des zéros, car ils semblent être le produit de 1, dans l'expression <1, par le facteur de conversion

des concentrations en unités par 100 mL, suivi d'un arrondissement au nombre entier le plus rapproché. On peut voir à la figure 67, en tenant compte des valeurs nulles, que les valeurs des missions de 1969 sont semblables à celles des missions de 1967 et 1968, bien que les données de la mission 110 de 1969 contiennent moins de valeurs nulles comparativement aux faibles valeurs obtenues lors des autres missions de ce groupe.

Les valeurs supérieures à 100 coliformes par 100 mL abondent pour les missions de 1966 et 1970. Les données de 1966 présentent le même pic de valeurs nulles que celles de 1967 et 1968, mais, comme on peut le voir au tableau 51, elles présentent une distribution particulière quand on les compare aux données de la mission 102 de 1968 et de la mission 107 de 1970. Le caractère bimodal des données transformées est donc dû à la gamme des valeurs de 0 à 100 environ qui présentent une distribution semblable à celle des missions de 1967 et 1968, tandis que la distribution du reste des valeurs est semblable à celle des valeurs de la mission de 1970.

Il faut conclure de tout ceci que les données de 1967, 1968 et 1969 peuvent probablement faire l'objet d'une étude des différences entre les années, mais que les comparaisons portant sur les années 1966 à 1970 ne peuvent être que qualitatives. Les différences entre les données de 1966 et 1970 et celles de 1967, 1968 et 1969 ne peuvent s'expliquer par des plans d'échantillonnage différents; elles pourraient s'expliquer par l'utilisation de milieux et de filtres différents lors de l'analyse (B.J. Dutka, communication personnelle).

On a montré précédemment que des distributions binomiales négatives décrivaient bien les concentrations de coliformes totaux des missions de 1968. Comme les données des missions de 1967 et 1969 comportaient un nombre important de valeurs nulles, les distributions lognormale et de Poisson ne pouvaient être utilisées; on a donc testé

la distribution binomiale négative qui, elle non plus, ne convenait pas aux données. On peut voir, à l'examen des trois missions de 1967 citées en exemples au tableau 54a, que, dans chaque cas, le nombre de valeurs correspondant à la partie de droite de la distribution excédait de beaucoup le nombre prévu pour une distribution binomiale négative. La gamme des valeurs des missions de 1968 est de 0 à 260, tandis qu'elle est de 0 à 150 000 pour les missions de 1967 et 1969. De plus, la valeur de 260 formait un groupe séparé suite au regroupement en grappes des données de la mission 108 de 1968. Il apparaît clairement que ces valeurs extrêmement élevées n'appartiennent pas à la même distribution que les autres. On a, pour exclure les valeurs modérément élevées, comparé les données de 1967 et 1969 à celles de 1968. On a exclu les valeurs supérieures à 200 quand leur exclusion permettait d'obtenir des groupes comparables à ceux de 1968 après regroupement. Cette méthode ad hoc a été considérée appropriée, car on ne sait pas si les concentrations de coliformes totaux des grands lacs seraient du type du groupe 1967-1968-1969 ou de celui du groupe 1966-1970. Il serait cependant justifié d'utiliser une méthode plus formelle si les données obtenues ultérieurement étaient du type du premier groupe.

On trouvera, au tableau 55, les résultats de ces tests d'ajustement de la distribution binomiale négative aux données des missions dont on a exclu les valeurs élevées. Les tests ont été effectués à l'aide des estimateurs par la méthode des moments et du maximum de vraisemblance de p et k, les paramètres de la distribution binomiale négative, les valeurs exclues et les estimations de p et k sont présentées au tableau 56 et l'on trouvera, au tableau 54b, le détail des tests relatifs aux missions 101, 103 et 105 de 1967. Les estimations du maximum de vraisemblance de p et k ont servi aux tests présentés au tableau 54b.

La distribution binomiale négative décrit beaucoup mieux les données des missions 101, 103 et 105 de 1967 lorsqu'on exclut les

valeurs élevées (voir les tableaux 54a et 54b). Il y a encore manque d'ajustement pour les données de la mission 101, mais cela s'explique par un nombre plus élevé que prévu de concentrations égales à 1 et non par un manque d'ajustement dans la partie supérieure de la distribution, comme cela avait été observé lors de l'utilisation de toutes les données. Les différents termes de la statistique $\chi^2$ de qualité de l'ajustement des missions 103 et 105 ne sont pas si élevés, pour toute valeur donnée, et l'ajustement est en général satisfaisant dans les deux cas. Les distributions de fréquences observées pour ces deux missions sont cependant quelque peu irrégulières, ceci en dépit du nombre important d'observations, n=161 et n=165.

L'examen des données des autres missions de 1967 et 1969 présentées au tableau 56 permet de noter que la distribution binomiale négative, basée sur les estimations par maximum de vraisemblance pour p et k, s'ajuste adéquatement aux données de toutes les missions, sauf la mission 109 de 1967. Comme dans le cas de la mission 101, ce phénomène s'explique en grande partie par un nombre de concentrations unitaires supérieur au nombre prévu. Dans le cas des missions 101 à 115 de 1967, des quatre missions de 1968 et de la mission 103 de 1969, les estimations de p et k obtenues par la méthode des moments diffèrent de façon marquée de celles obtenues par la méthode du maximum de vraisemblance, et pour chacune de ces missions, à l'exception de la mission 109 de 1967 où les deux séries d'estimations donnent des résultats fortement significatifs, l'interprétation de l'ajustement serait différente selon le couple d'estimation de p et k utilisé. A l'exception de la mission 101 de 1967, l'ajustement est toujours moins bon lorsqu'on utilise les estimations par la méthode des moments. Comme ces dernières ne sont pas valables pour des valeurs de k aussi faibles, la conclusion basée sur les données de 1968, voulant que l'on utilise les estimateurs de maximum de vraisemblance, s'en trouve renforcée.

Les distributions de fréquences observées indiquent que les distributions de 1967, 1968 et 1969 sont très fortement biaisées. De plus, on a montré en ajustant une distribution théorique aux données que, à deux exceptions près, une distribution binomiale négative convenait aux données d'une même mission. Ce résultat est remarquable étant donné la grande variation du nombre des stations d'échantillonnage et le nombre élevé de valeurs nulles observées (tableau 56). Quand on peut, en modifiant les paramètres, ajuster un même modèle à une série d'ensembles de données, telles les données des missions utilisées ici, ces ensembles peuvent être comparés à l'aide des valeurs des paramètres. Les estimations par maximum de vraisemblance de p et k ont été reportées à la figure 68 pour les années 1967, 1968 et 1969. La variation erratique des paramètres semble cependant être la seule certitude à laquelle il soit possible de conclure. L'examen de toutes les données d'une même mission ne devrait pas renseigner davantage vu la variabilité du nombre et de l'emplacement des stations d'échantillonnage.

Comme on peut le voir au tableau 57, la distribution lognormale s'ajuste bien aux données de la mission 107 de 1970 et il semble (voir plus haut) que l'ajustement aux données de la mission 111 de 1966 soit plus précis si l'on effectue une transformation logarithmique, comme pour les données de 1970. Mais les valeurs de 23 des 92 observations sont nulles et ne peuvent subir une transformation directe. On a ajouté un à ces valeurs; la transformation effectuée avant le test d'ajustement a donc été log(R+1). Comme on peut le voir au tableau 57, la distribution lognormale ne s'ajuste pas à ces données à cause du nombre important de valeurs nulles et de la présence de deux autres pics dans la distribution des fréquences observées. Le modèle appelé distribution Δ (Aitchison et Brown, 1957) qui suppose que la population comprend une proportion δ de valeurs nulles et que les valeurs non nulles ont une distribution lognormale, serait plus approprié. Les calculs supplémentaires n'ont pas été effectués, car il a été établi de façon

suffisante que les données de 1966 différaient de celles de toutes les autres années.

La méthode de regroupement des probabilités, où le nombre de groupes est fixé à 4 et où les valeurs élevées sont exclues, a été appliquée aux 16 missions; les résultats obtenus sont présentés au tableau 58 et à la figure 69. Les valeurs exclues ont été placées dans la zone 5 et sont présentées, pour chaque mission, à la onzième colonne du tableau 58. La structure saisonnière décrite pour les missions de 1968 est, du moins partiellement, confirmée pour les autres années.

Les premières missions sont la mission 101, du 30 mai au 8 juin 1967, la mission 102, du 17 au 22 mai 1968, et la mission 103, du 30 mai au 4 juin 1969. Dans les trois cas, la plupart des stations du large se trouvent dans la zone 1; les autres se situent dans une zone près de Cleveland et dans des groupes discontinus le long de la rive sud. En 1968 et 1969, la zone 2 se situe près de la rive à l'extrémité nord-est du lac et les zones de concentrations élevées se situent dans le bassin ouest mais, en 1968, on les retrouve aussi dans la partie nord-ouest du bassin central. A noter que les moyennes pour les zones correspondantes sont comparables: 0.10, 0.44 et 0.19 pour la zone 1; 7.50, 7.14 et 10.31 pour la zone 2; 40.00, 43.33 et 51.40 pour la zone 3 et 87.50, 79.50 et 130 pour la zone 4, ceci dans l'ordre des missions 101 en 1967, 102 en 1968 et 103 en 1969.

Les zones de concentrations élevées, lors de la deuxième mission de 1968 effectuée du 29 juillet au 3 août, s'étaient étendues à l'extrémité est du lac. Le phénomène n'a pu être confirmé dans le bassin central à cause d'un nombre insuffisant d'échantillons. Le nombre de missions a été plus élevé en 1967 et la deuxième a été effectuée plus tôt. Pendant la mission 103, du 20 au 29 juin, les zones de concentrations élevées traversaient le lac. De fait, on

note pour les missions 103, 105, 107, 113 et 115 de 1967 des bandes de concentrations élevées traversant le lac de la rive nord à la rive sud en divers endroits. Les missions 109, du 21 au 30 août, et 111, du 11 au 17 septembre, n'ont pas permis de noter l'existence de telles bandes, mais l'échantillonnage du bassin central a été moins dense lors de ces missions. Les concentrations les plus élevées de 1967 ont été notées lors de la mission 103, du 20 au 29 juin, et celles de 1968 lors de la mission 108, du 29 juillet au 3 août. Aucune mission n'ayant été effectuée à la fin de juin 1968, il est impossible de savoir si les concentrations maximales sont apparues plus tard ou si elles n'ont pas été notées. Les concentrations de 1966 et 1970 sont très différentes de celles de 1967, 1968 et 1969, mais leurs répartitions spatiales sont qualitativement semblables à celles notées pendant la même période de l'année en 1967, 1968 et 1969. La mission 111 effectuée du 8 au 14 août 1966 et la mission 107 effectuée du 28 juillet au 2 août 1970 sont celles où l'on note des bandes de concentrations très élevées traversant le lac.

Les dernières missions, effectuées en octobre ou novembre de 1967, 1968 et 1969, semblent révéler les mêmes répartitions qu'au printemps, c'est-à-dire moins de stations dans les zones de concentrations élevées du centre du lac. On note cependant, pour la mission d'octobre 1967, la présence au travers du lac d'une bande passée inaperçue en août et en septembre.

La répartition spatiale a donc été caractérisée qualitativement pour trois périodes de l'année. Les résultats des missions sont quantitatifs, mais il est impossible de décrire quantitativement les zones applicables aux données de toutes les années à cause des différences entre les données d'une année à l'autre, du nombre restreint de missions effectuées à chaque année, à l'exception de 1967, et de l'irrégularité du plan d'échantillonnage. La moyenne des zones 1, 2, 3 et 4 ainsi que celles de la zone 1 et de la zone 2 peuvent être composées pour 1967, 1968 et 1969 à la partie

supérieure de la figure 68. Les concentrations maximales de la mission 101 de 1967 et de la mission 108 de 1968 ressortent dans la courbe des moyennes générales ainsi que dans les zones 1 et 2 de 1967, et dans la seule zone 1 de 1968. Les régions discontinues de concentrations élevées le long des rives sont ressorties lors de chaque mission, mais non nécessairement aux mêmes endroits, sauf la région de Cleveland où des concentrations élevées ont toujours été observées à certaines stations et celle d'Erié où de telles concentrations sont presque toujours notées. Pour mieux comparer les zones au printemps et en été, on a reporté à la figure 70 les stations de la zone 2 lors de la mission 101 du 30 mai au 8 juin et lors de la mission 103 du 20 au 29 juin 1967. Les emplacements des stations des zones 3, 4 et 5 des missions de 1967 sont présentés à la figure 71. De plus, les stations exclues des calculs des zones 1 à 4, regroupées ici dans la zone 5, se retrouvent, à quatre exceptions près, dans les environs de Cleveland, de Toledo, d'Erié, de l'embouchure de la rivière Détroit ou de l'embouchure de la rivière Grand en Ontario (voir tableau 59). Les quatre exceptions sont aussi présentées au tableau et peuvent être situées sur les cartes de la figure 69. Il faut noter qu'entre 1967 et 1969, les valeurs supérieures à 1000 coliformes par 100 mL n'ont été notées que dans les environs de Toledo, de Cleveland et de l'embouchure de la rivière Grand. Ces bandes de concentrations élevées (généralement les zones 2 et 3) qui traversent le lac apparaissent probablement sous l'effet des courants, car il faut qu'il y ait transport à partir de la rive. Ce phénomène peut tenir à l'une ou l'autre des deux caractéristiques suivantes de la circulation dans le bassin central (Hamblin, 1971; Simons, 1976): (1) des cellules de circulation issues des courants dus à l'écoulement qui longent la rive nord gagnent l'ensemble du bassin ou (2) la plupart des courants de surface d'été se dirigent au large de la rive nord.

## ANALYSE DES DONNEES SUR LES COLIFORMES FECAUX

On dispose de données sur les coliformes fécaux pour 15 missions effectuées entre 1967 et 1975, mais les concentrations de coliformes totaux n'ont été déterminées que lors de 11 missions (tableau 50). Plusieurs raisons font que la relation entre les concentrations de coliformes totaux et fécaux présente un intérêt; on mettra donc l'accent, dans la présente section, sur le rapport des coliformes plutôt que sur leurs concentrations. Santé et Bien-être social Canada (1979) recommande, dans ses lignes directrices sur le traitement des eaux brutes utilisées comme source d'eau de consommation, que trois critères soient respectés et l'on trouve, dans chacun de ces critères, mention de concentrations maximales permises de coliformes tant fécaux que totaux. Les lignes directrices sur les eaux potables épurées stipulent aussi qu'elles ne doivent contenir aucun coliforme fécal. On s'intéresse, dans ce contexte, à la distribution des fréquences empiriques des coliformes fécaux en fonction de la concentration des coliformes totaux. On cherche aussi à savoir si la concentration de coliformes totaux est un indicateur approprié des agents pathogènes présents, les coliformes totaux regroupent des organismes d'origine tant fécale que non fécale. On calcule donc, à cet effet, le rapport dont on a déjà traité.

La distribution des fréquences des concentrations de coliformes fécaux notées lors des 15 missions présente une proportion élevée de valeurs nulles (0.52 à 0.87). Le reste des observations s'étale sur une large gamme et, dans la plupart des cas, la fréquence de ces observations est égale à un (tableau 60). Des concentrations de coliformes fécaux plus élevées ont été notées lors des missions de 1967. Les emplacements des stations où les concentrations de coliformes fécaux étaient supérieures à 200 sont présentés dans le bas du tableau 60. A quelques exceptions près, comme dans le cas des coliformes totaux, les concentrations élevées apparaissent dans les

environs de Cleveland, d'Erié, de Toledo et de la rivière Grand. Les valeurs observées lors de la mission 110 de 1969 et inférieures à un certain seuil ont été notées comme nulles pour les mêmes raisons que celles déjà données pour les coliformes totaux.

Comme première étape dans l'étude de la relation entre les concentrations de coliformes totaux et fécaux, on a reporté graphiquement les concentrations de coliformes fécaux en fonction de celles des coliformes totaux aux stations de chacune des missions de 1967 et 1968 pour les concentrations de coliformes totaux supérieures à 100 bactéries par 100 mL (graphique du haut de la figure 72). Les gammes de valeurs sont de 0 à 51 000 pour les coliformes totaux et de 0 à 12 000 pour les coliformes fécaux. On trouvera, à la figure 72, le même graphique pour les concentrations de coliformes totaux inférieures à 1000. Les concentrations élevées non indiquées sur les graphiques sont présentées plus bas; la première rangée est celle des coliformes totaux et, la deuxième, celle des coliformes fécaux.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1000 | 2600 | 6000 | 6300 | 8300 | 11 000 | 12 000 | 22 000 | 51 000 |
| 400 | 340 | 880 | 5000 | 400 | 1 600 | 3 500 | 250 | 12 000 |

La seule conclusion que l'on puisse tirer de ces graphiques est l'absence de relation pour les faibles concentrations de coliformes totaux, de 0 à 30, et que, même si les concentrations de coliformes fécaux élevées correspondent à des concentrations de coliformes totaux élevées, la variance est très grande.

Le critère de qualité des eaux potables donné plus haut, à savoir qu'aucun coliforme décelé ne soit fécal, appelle une distribution de probabilités conditionnelles si on veut l'exprimer en termes de probabilités. Des concentrations de coliformes totaux et fécaux de mêmes échantillons étant disponibles, il est possible d'étudier la distribution des fréquences observées des coliformes fécaux qui est conditionnelle pour une concentration de coliformes

totaux donnée. On procède en obtenant la distribution des fréquences des concentrations des coliformes fécaux pour lesquelles les concentrations de coliformes totaux sont les mêmes. Soit $R_t$ la concentration de coliformes totaux et $R_f$ la concentration de coliformes fécaux, la fonction de fréquences conditionnelle de $R_f$, pour $R_t$ donné, $f(r_f/R_t=r_t)$, est donnée par le nombre de stations où la concentration de coliformes fécaux est égale à $r_f$ et où la concentration de coliformes totaux est égale à $r_t$ divisé par le nombre de stations où la concentration de coliformes totaux est égale à $r_t$. Ces distributions de fréquences sont présentées, à la figure 73, pour les missions 103, 105 et 107 de 1967 et 102, 108 et 112 de 1968, en fonction des concentrations de coliformes totaux allant de 0 à 20 et, au tableau 61, pour toutes les données de 1967 et 1968. Il devient apparent qu'il faut disposer d'une grande quantité de données pour déterminer l'allure de la fonction de fréquences conditionnelles. Comme on peut le voir au tableau 61, les concentrations de coliformes fécaux sont faibles quand celles des coliformes totaux sont élevées. Nous ne disposons de nombreuses observations que pour les faibles concentrations de coliformes totaux, de sorte qu'il est presque impossible d'interpréter $f(R_f/R_t)$ pour les valeurs supérieures à $R_t > 40$. Pour les faibles valeurs de $R_t$, $f(R_f/R_t)$ est biaisée, la valeur modale se situant à 0; la fonction devient cependant moins biaisée à mesure que $R_t$ augmente. Cette dernière observation est caractéristique de plusieurs distributions biaisées. Ainsi, à mesure que la moyenne de la distribution de Poisson augmente, la fonction de distribution devient de plus en plus symétrique. On n'a pas déterminé ici l'allure de la distribution conditionnelle, mais la présente discussion est importante car elle constitue la première étape de l'étude de lignes directrices du genre de celles présentées plus haut. L'importance de connaître l'allure de la distribution de fréquences lors de l'élaboration de règlements et de lignes directrices a été traitée par Esterby (1982).

Il faut aussi noter que, aux faibles concentrations, la concentration de coliformes fécaux est parfois supérieure à celle des coliformes totaux; ce résultat est plausible lorsqu'on effectue des analyses indépendantes portant sur un même échantillon, mais non si les coliformes fécaux sont dénombrés sur la plaque des coliformes totaux. Cette question, dont on a déjà traité, relève de l'hypothèse relative à l'estimation du rapport des coliformes.

Les zones déterminées à la section précédente à l'aide de toutes les concentrations de coliformes totaux ont servi à l'étude de la distribution spatiale des coliformes fécaux. Des coliformes fécaux n'ayant pas été dénombrés à toutes les stations où l'on a relevé des coliformes totaux, un sous-ensemble de stations communes a été identifié pour chaque zone. Les concentrations de coliformes fécaux aux stations communes sont présentées sous forme résumée au tableau 62, et celles des coliformes totaux, au tableau 63. L'étude de la relation entre les concentrations de coliformes totaux et celles des coliformes fécaux, présentée au début de la présente section, a été basée sur des regroupements de séries de données. La répartition en zones permet d'étudier les données après élimination de la composante spatiale.

On a reporté à la figure 74 les concentrations de coliformes fécaux en fonction de celles des coliformes totaux pour chacune des zones des missions de 1967 et 1968, et à la figure 75 les concentrations moyennes de coliformes fécaux en fonction de celles des coliformes totaux. A la figure 74, la concentration des coliformes fécaux est reportée en ordonnée et celle des coliformes totaux en abscisse; la fréquence de chaque couple de valeurs est indiqué en lieu du symbole (•) si ce nombre est supérieur à 1. Ici aussi, l'absence de relation aux faibles concentrations est apparente, si l'on fait exception de la zone 2 de la mission 112 de 1968.

Comme cela a été mentionné, le rapport, $\rho$, de la concentration de coliformes fécaux à celle des coliformes totaux peut être estimé à partir du rapport des moyennes correspondantes des échantillons. On a procédé de la sorte en n'utilisant que les stations communes aux zones 1, 2 et 3 des missions de 1967, 1968 et 1969 et en se servant des moyennes des coliformes totaux pour l'ensemble des données. Les résultats sont présentés au tableau 64 et les rapports $\rho$ estimés aux stations communes sont présentés graphiquement à la figure 76.

Le rapport de la zone 1 varie de façon erratique pendant les missions de 1967; ceci surtout à cause de concentrations de coliformes fécaux supérieures aux concentrations de coliformes totaux dans au moins une des stations de cette zone. Comme on peut le voir à la figure 74, cela s'est produit lors des missions 105, 107, 111, 113 et 115 où les rapports de la zone 1 sont plus élevés que ceux des missions 103 et 109. Une droite à 45 degrés a été tracée sur les graphiques de la figure 74 dans lesquels les concentrations de coliformes fécaux sont supérieures aux concentrations correspondantes de coliformes totaux à une ou plusieurs stations. La station de la zone 1 de la mission 111 où la concentration de coliformes fécaux était de 40 a été exclue des calculs. Dans la zone 2, la concentration de coliformes fécaux a été supérieure à celle des coliformes totaux à trois reprises (missions 105, 113 et 115) et ici aussi, comme pour la zone 1, ce sont les missions présentant un rapport élevé. On peut se demander si ces stations devraient être exclues des calculs. Comme on peut le voir au tableau 60, les distributions de fréquences des coliformes fécaux sont fortement biaisées. Les distributions de fréquences des concentrations de coliformes fécaux de la zone 2 des missions de 1967 sont aussi fortement biaisées, comme le sont d'ailleurs les distributions de fréquences conditionnelles (tableau 61). La probabilité d'obtenir une concentration de coliformes fécaux relativement élevée dans une zone n'est pas négligeable; ces stations n'ont donc pas été exclues.

Les rapports obtenus pour les missions 105, 113 et 115 ($\hat{\rho} > 0.20$) sont supérieurs à ceux obtenus pour les missions 103, 107, 109 et 111 ($\hat{\rho} < 0.10$), car quelques concentrations de coliformes fécaux élevées ont été notées lors des premières missions.

En 1968, le rapport de la mission 108 est inférieur à ceux des missions 102 et 112. On ne peut tirer de conclusion pour l'année 1969, le rapport ne pouvant être estimé que pour une seule mission. Les variations des rapports des années 1967 et 1968 présentent une même caractéristique. Ces rapports sont faibles pendant les missions où les concentrations de coliformes totaux sont maximales et où les zones de concentration élevée s'étalent au travers du lac. En 1968, cette situation ne s'est produite que lors de la mission 108; il y a donc une différence de distribution spatiale qui concorde avec celle des rapports. Cependant, en 1967, la zone 2 ne forme pas de bandes à travers le lac lors de la mission 101 et les missions 109 et 111, en août et septembre, mais elle n'est pas absente du centre du lac lors des missions de début et de fin d'année comme en 1968. De plus, les différences entre les rapports des missions de 1968 sont accompagnées d'une différence dans la relation entre les concentrations de coliformes fécaux et totaux, comme cela est indiqué à la figure 74. Il semble y avoir absence de relation entre les concentrations de coliformes fécaux et totaux pendant la mission 102 de 1968; on note que la variance des concentrations de coliformes fécaux augmente avec celles des coliformes totaux pour la mission 108 et que les concentrations de coliformes fécaux augmentent avec celles des coliformes totaux pour la mission 112. Cela va nettement à l'encontre de la cause des différences entre les missions de 1967.

Ces écarts peuvent s'expliquer par les différences dans les plans d'échantillonnage des coliformes fécaux de 1967 et de 1968 (figure 77). En 1967, les échantillons ont été prélevés dans le bassin ouest ainsi que dans les environs de Cleveland, d'Erié et de l'embouchure de la rivière Grand en Ontario, malgré une variation

considérable d'une mission à l'autre. En 1968, les échantillons ont été prélevés le long de la plus grande partie de la rive, mais, ici encore, la variation d'une mission à l'autre est considérable. Les concentrations de coliformes fécaux de 1967 ont donc été notées à des stations réputées proches de sources de pollution bactérienne terrestres, tandis que les concentrations de 1968 ont été mesurées à des stations situées près de la rive, à proximité ou loin de sources de pollution terrestres. A noter que seule la mission de 1969 permet la comparaison des concentrations de coliformes fécaux et totaux pour tout le lac. Il faudra donc attendre de disposer de données supplémentaires avant de pouvoir tirer une conclusion quant à la structure saisonnière des rapports.

La gamme des valeurs plausibles de $\rho$ peut être obtenue à partir de la fonction de vraisemblance conditionnelle relative de $\rho$. On peut déduire de la fonction de vraisemblance conditionnelle $L_c(\rho)$ présentée précédemment que:

$$L_c(\rho) \propto \rho^{f.} / (1+\rho)^{t.+f.}$$

où pour n couples de concentrations de coliformes fécaux et totaux indépendantes $(f_1,t_1),(f_2,t_2),\ldots,(f_n,t_n)$ et $f. = \sum_{i=1}^{n} f_i$ et $t. = \sum_{i=1}^{n} t_i$. La fonction de vraisemblance relative est donnée par

$$R_c(\rho) = \frac{L_c(\rho)}{\max_{\rho} L_c(\rho)}$$

Cette fonction a été reportée à la figure 78 pour la zone 2 des missions 103 et 105 de 1967 et 102 et 108 de 1968. Il y a chevauchement de la gamme des valeurs plausibles, $R_c(\rho) \geq 0.50$, pour les quatre missions présentées à la figure 76; cela signifie que les rapports ne diffèrent pas vraiment. A nouveau, la seule conclusion logique que l'on puisse tirer est que les données sont

insuffisante pour déterminer la fonction $R_c(\rho)$. Il est impossible de dire si l'hypothèse voulant que les concentrations de coliformes fécaux et totaux présentent une distribution de Poisson est valable.

Des exemples de cartes disponibles pour les concentrations de coliformes fécaux sont présentés à la figure 79. La concentration de coliformes fécaux, pour les zones de coliformes totaux, est représentée par un point (.) lorsqu'elle est nulle et par sa valeur lorsque celle-ci est supérieure à 0. On peut voir, pour les missions 103 et 105 de 1967, que les conclusions relatives aux concentrations de coliformes fécaux ne visent que les environs des régions urbaines ou des embouchures de cours d'eau où, comme cela a été montré plus haut, les concentrations de coliformes totaux sont élevées. Aucune déduction relative au rapport des concentrations ne peut donc être faite pour l'ensemble du lac.

## RESUME ET DISCUSSION

Les distributions de fréquences des concentrations de coliformes totaux observées lors de chacune des missions montrent que les données de 1966 et 1970 diffèrent nettement de celles de 1967, 1968 et 1969. Pour ces trois dernières années, il y a un nombre important de concentrations nulles et la distribution pour chaque mission est fortement biaisée. La plupart des concentrations se situent dans la gamme de 0 à environ 200 organismes par 100 mL; une distribution binomiale négative peut être ajustée aux données de deux des 14 missions pour cette gamme de concentrations. Le manque d'ajustement aux données de ces deux missions s'explique par un pic plus large dans la distribution observée par rapport à la distribution binomiale négative ajustéee. On note, pour chacune des missions de 1967 et 1969, quelques concentrations élevées qui n'appartiennent pas à la même population que les autres et qui correspondent à des stations situées près d'importantes zones urbaines ou des embouchures de cours d'eau. Les données de 1970 sont

complètement différentes. Les concentrations de coliformes totaux varient de 20 à 2700, et il est possible d'y ajuster facilement une distribution lognormale. Les données de 1966 semblent présenter les caractéristiques des deux autres types, mais il est impossible de leur ajuster une distribution lognormale ou binomiale négative.

Les concentrations de la gamme réduite, 0 à 200, présentent toujours une variation importante, ceci pour toutes les missions. Cette variation doit s'expliquer par des différences spatiales, le temps consacré à chaque mission étant limité. Cela a été confirmé par les résultats du regroupement par probabilités appliqué à toutes les missions de 1966 à 1970. En dépit de différences entre les années, une tendance saisonnière ressort dans la distribution spatiale. Au début de l'année, la zone des concentrations les plus faibles couvre la plus grande partie du lac, les zones de concentrations plus élevées formant des régions discontinues le long des rives. Plus tard, les concentrations atteignent des maximums et les zones 2, et parfois 3, s'étendent en bandes au travers du lac. Les régions discontinues de concentrations plus élevées le long des rives sont toujours présentes et, de fait, on les retrouve à chaque mission, même si leur position n'est pas nécessairement conservée. A l'automne, l'allure générale semble redevenir celle de la distribution spatiale antérieure, mais on note des différences entre les années qui sont surtout dues à la présence de bandes de zones 2 perpendiculaires à la rive vers la fin de 1967. L'existence des régions discontinues le long des rives peut être attribuée à des sources riveraines, ce qui confirme les prévisions. Les bandes traversant le lac doivent cependant s'expliquer par certains modes de circulation au sein du lac. L'étendue des zones 2 et 3 sur tout le lac lors de la mission de juillet et août 1968 avait été signalée par El-Shaarawi et coll. (1981). Ce n'est qu'après analyse de toutes les données disponibles qu'il est apparu que cette tendance se répétait, la bande pouvant cependant se retrouver en des endroits différents.

Reitz (1973) donne les concentrations de coliformes totaux déterminées aux stations de prise d'eau de la rive sud du lac Erié. Il signale que ces concentrations, mesurées à des distances considérables de la rive, sont plus faibles que celles notées sur la plage, mais supérieures à la moyenne de deux coliformes par 100 mL obtenue à environ deux milles de la rive. Il a attribué ces fortes concentrations à des processus riverains locaux et aux courants de direction est longeant la rive sud. Rao et Burnison (1976) rapportent les mêmes données sur les coliformes totaux de 1970 que celles analysées ici et signalent que certaines régions, comme celles de Cleveland, d'Erié et de la partie nord-ouest du bassin central, présentent des concentrations plus élevées (>500) que le reste du lac. Ils attribuent aussi la valeur élevée de ces concentrations dans la partie nord-ouest du bassin central à certains modes de circulation. Les observations faites au présent chapitre concordent donc avec celles faites par ces auteurs. La description plus détaillée de l'évolution spatiale et saisonnière du lac Erié faite au présent chapitre fournit cependant beaucoup plus d'informations que l'on en disposait antérieurement. De plus, les zones de concentrations élevées qui s'étendent sous forme de bandes à travers le lac, et qui ont été attribuées à des courants au sein du lac, sont nouvelles car elles doivent avoir pour origine des cellules de circulation et non simplement des courants riverains. Le moins que l'on puisse dire est que ces bandes donnent lieu à des concentrations plus élevées que prévues au large.

Les températures chaudes de l'eau se traduisent par un taux plus élevé de survie des bactéries coliformes (Reitz, 1973). On a, pour savoir si la température était liée à la présence de ces bandes, reporté graphiquement les concentrations de coliformes totaux et les températures dans la région d'une bande de stations de zone 2 lors des missions 103 et 107, et l'on a procédé de la même façon avec les valeurs de la mission 109 en utilisant la région de la mission 103, toutes ces missions ayant été effectuées en 1967. Il faut noter que

la zone 2 était absente du milieu du lac lors de la mission 109. On peut voir que des températures chaudes ne sont pas à l'origine des bandes qui étaient dues soit à une température généralement plus élevée du lac ou à des températures plus chaudes dans la bande de la zone 2, comparativement aux stations avoisinantes. La température du lac était plus élevée lors des missions 107 et 109 que lors de la mission 103 et l'on note la présence de bandes lors des missions 103 et 107, mais non lors de la mission 109. Les écarts de température dans les régions, notés au cours des trois missions, forment un gradient de la rive nord à la rive sud sans qu'il n'y ait de différence entre les stations de la zone 1 et celles des zones 2 et 3.

Les données ne permettent pas de comparer les années à cause de différences inhérentes dans les concentrations, de l'irrégularité du plan d'échantillonnage et du nombre insuffisant de missions la plupart des années. Cela s'applique aussi aux concentrations de coliformes fécaux. Il faut aussi souligner que les concentrations de coliformes totaux et fécaux enregistrées en 1969 comme inférieures à une certaine valeur ont été converties en valeurs nulles afin d'appliquer la méthode d'enregistrement des données utilisée lors des autres années.

Les concentrations de coliformes fécaux d'une mission consistent surtout en valeurs nulles et quelques valeurs non nulles très variables. Dans ce cas aussi, des concentrations élevées pouvant atteindre 12 000 organismes par 100 mL ont été notées dans les environs de zones urbaines importantes ou d'embouchures de cours d'eau et le manque d'uniformité dans le plan d'échantillonnage est responsable de l'importante variation de la concentration maximale de coliformes fécaux notée lors de diverses missions, les valeurs des régions à concentration élevées ayant parfois été incluses.

La relation entre les concentrations de coliformes fécaux et totaux a été abordée de diverses façons. Des graphiques de toutes

les concentrations de coliformes fécaux et totaux, obtenues à la même station et à l'intérieur des quatre zones à partir de l'ensemble des données sur les coliformes totaux, ont montré que, pour les concentrations de coliformes totaux inférieures à 30, il n'y a pas de relation entre les concentrations de coliformes fécaux et totaux, ceci à l'exception de la zone 2 de la mission 112 de 1968. Si l'on utilise une gamme plus large, les concentrations de coliformes fécaux sont élevées, mais la variation est énorme. L'examen de cette relation basée sur le rapport de la concentration des coliformes fécaux à la concentration moyenne des coliformes totaux par zone ne permet de tirer qu'une seule conclusion, à savoir que le rapport est faible pour une concentration de coliformes totaux maximale, et les données ne permettent pas de tirer aucune conclusion quant aux effets saisonniers. La distribution des fréquences conditionnelles des concentrations de coliformes totaux permet de tirer une conclusion préliminaire, à savoir que cette distribution, comme la distribution non conditionnelle, est biaisée vers la droite. La signification de ce phénomène, dans le cadre de l'interprétation des lignes directrices sur la qualité de l'eau, est que, pour une concentration donnée de coliformes totaux, la probabilité d'obtenir une concentration de coliformes fécaux non nulle est plus importante que celle que l'on obtiendrait avec une distribution symétrique.

L'utilisation d'un organisme indicateur ayant pour but de déceler des organismes pathogènes, on peut étudier la validité des coliformes totaux comme indicateur de contamination fécale en comparant les concentrations de coliformes fécaux et totaux aux mêmes stations. La fonction des fréquences conditionnelles s'avère ici plus utile que le rapport des concentrations, car le nombre de fois qu'une concentration de coliformes fécaux nulle est observée pour une concentration de coliformes totaux non nulle donnée permet de savoir quand la concentration de coliformes totaux permet de déceler des organismes d'origine strictement non fécale. Lors des missions de 1967 et 1968, où seulement des stations situées le long de la rive ou

dans les environs d'un centre urbain important ou d'une embouchure de cours d'eau ont fait l'objet d'un échantillonnage. 129 des 215 stations où la concentration de coliformes totaux n'était pas nulle présentaient une concentration de coliformes fécaux nulle. On peut voir au tableau 61 que les concentrations de coliformes totaux étaient inférieures à 50 dans 126 des 129 stations. Ce phénomène peut être étudié de façon plus détaillée à la figure 74. Ainsi, dans les zones 1, 2, 3 et 4 de la mission 103 de 1967, la proportion de stations où la concentration de coliformes fécaux était nulle lorsque celle des coliformes totaux ne l'était pas est respectivement, de 10/11 = (0.91), 9/13 = (0.69), 3/5 = (0.60) et 0/1 = (0). L'organisme indicateur donne moins de faux résultats positifs aux concentrations élevées.

Une autre façon d'évaluer la proportion de coliformes non fécaux dans les coliformes totaux consiste à comparer la turbidité et la concentration de coliformes totaux. La turbidité est plus élevée dans le bassin ouest à cause de l'important apport de ses tributaires en limon. Ce dernier regorge de bactéries courantes du sol qui entrent dans le dénombrement des coliformes totaux.

On a, pour les missions de 1968, reporté graphiquement la concentration des coliformes totaux en fonction de la turbidité aux stations du bassin ouest (figure 80). Ces données, très limitées, n'indiquent pas d'augmentation des concentrations de coliformes totaux en fonction de la turbidité.

Toutes les données disponibles pour 1968 ont aussi fait l'objet d'un examen. La concentration de coliformes totaux a été reportée à la figure 81 en fonction de la turbidité, ceci pour chaque mission. On peut voir à cette figure ainsi qu'au tableau 65 et à la figure 82 où les valeurs moyennes de la turbidité et de la concentration de coliformes totaux sont présentées selon les zones définies à partir de l'ensemble des concentrations de coliformes

totaux, qu'il n'y a pas de relation entre la concentration de coliformes totaux et la turbidité aux faibles valeurs de turbidité (<4). Une telle relation pourrait apparaître pour une gamme plus étendue, mais le nombre de valeurs de turbidité élevée notées en 1968 est insuffisant pour permettre de tirer des conclusions. Il serait intéressant de poursuivre cette étude en incluant un facteur pour isoler la turbidité due au limon de celle due à d'autres sources et en s'intéressant aux bassins un à un. Une telle analyse, si on établissait une relation, devrait montrer que les bactéries du sol contribuent à la concentration de coliformes totaux quand celle des coliformes fécaux est nulle.

Tous les résultats présentés ci-haut découlent de l'application de méthodes statistiques dont le but est d'identifier et de quantifier les sources de variation. Des méthodes bien connues, ainsi que certaines méthodes relativement nouvelles, ont été utilisées. Le concept simple de la distribution des fréquences observées a permis d'isoler des séries de données très différentes même si elles ont, semble-t-il, été recueillies à l'aide des mêmes méthodes analytiques. L'ajustement de distributions de probabilité aux données a permis de mieux quantifier les résultats. La comparaison des estimations, établies par maximum de vraisemblance, des paramètres de la distribution binomiale négative avec celles de la méthode des moments a fait recommander l'utilisation de la première méthode pour certaines données, comme les concentrations de coliformes totaux des années 1967, 1968 et 1969. Ce résultat et la fonction de vraisemblance mixte pour ces paramètres ont une importance considérable lorsqu'on utilise la distribution binomiale négative.

La détection d'une tendance saisonnière de la distribution spatiale des concentrations de coliformes totaux, obtenue à l'aide des zones, illustre, en dépit du grand nombre de limitations inhérentes aux données, la puissance de la technique du regroupement

par probabilités. Il faut noter que de meilleures séries de données permettraient d'éliminer la limitation arbitraire à quatre zones; le nombre de zones choisies serait alors celui indiqué par la méthode. On a, dans le présent cas, considéré qu'il était plus important de pouvoir effectuer une comparaison entre les années. On présente, pour terminer, une estimation d'un rapport de concentrations bactériennes et des distributions de fréquences conditionnelles. Ces dernières se sont souvent avérées très utiles pour comparer des concentrations de coliformes fécaux et totaux.

Tableau 50. Résumé du nombre de stations où les concentrations de coliformes totaux ou fécaux, ou les deux, ont été déterminées, ceci pour toutes les données de surface disponibles*

| Année-mission | Dates | Nombre de stations | | |
|---|---|---|---|---|
| | | Coliformes totaux | Coliformes fécaux | Col. totaux et fécaux |
| 66-111 | 8-14 août | 92 | -- | -- |
| 67-101 | 30 mai - 8 juin | 150 | -- | -- |
| 67-103 | 20-29 juin | 166 | 47 | 47 |
| 67-105 | 10-19 juillet | 170 | 57 | 57 |
| 67-107 | 31 juillet - 10 août | 99 | 31 | 31 |
| 67-109 | 21-30 août | 97 | 30 | 30 |
| 67-111 | 11-17 sept. | 97 | 27 | 27 |
| 67-113 | 2-9 oct. | 95 | 33 | 30 |
| 67-115 | 24-30 oct. | 75 | 21 | 21 |
| 68-102 | 17-22 mai | 62 | 35 | 34 |
| 68-108 | 29 juillet-3 août | 62 | 35 | 33 |
| 68-109 | 31 août-2 sept. | 28 | -- | -- |
| 68-112 | 5-10 nov. | 52 | 30 | 23 |
| 69-103 | 30 mai-4 juin | 59 | -- | -- |
| 69-110 | 15-20 oct. | 73 | 72 | 72 |
| 70-107 | 28 juillet-2 août | 47 | -- | -- |
| 74-104 | 21-25 août | -- | 30 | -- |
| 75-101 | 3-10 avril | -- | 32 | -- |
| 75-106 | 24-29 juin | -- | 12 | -- |
| 75-107 | 5-10 août | -- | 31 | -- |

* Seules les concentrations des échantillons prélevés à 1 m de profondeur ont été utilisées, les concentrations n'ayant été déterminées systématiquement qu'à cette profondeur.

Tableau 51. Distributions des fréquences observées, f(r), des concentrations de coliformes totaux, r, lors des missions 111 de 1966, 102 de 1968 et 107 de 1970

| Mission 111 - 1966 | | Mission 102 - 1968 | | Mission 107 - 1970 | |
|---|---|---|---|---|---|
| r | f(r) | r | f(r) | r | f(r) |
| 0 | 23 | 0 | 29 | 20 | 1 |
| 1 | 4 | 1 | 9 | 62 | 1 |
| 3 | 4 | 2 | 5 | 80 | 1 |
| 5 | 3 | 3 | 2 | 82 | 1 |
| 0 | 1 | 4 | 2 | 90 | 1 |
| 9 | 1 | 5 | 3 | 110 | 2 |
| 10 | 14 | 6 | 1 | 130 | 2 |
| 14 | 2 | 7 | 1 | 140 | 1 |
| 17 | 1 | 9 | 1 | 180 | 1 |
| 20 | 3 | 10 | 1 | 200 | 2 |
| 25 | 1 | 11 | 1 | 210 | 1 |
| 30 | 4 | 12 | 1 | 220 | 1 |
| 50 | 5 | 18 | 1 | 230 | 1 |
| 60 | 1 | 35 | 1 | 260 | 1 |
| 70 | 1 | 36 | 1 | 300 | 2 |
| 100 | 1 | 59 | 1 | 310 | 2 |
| 110 | 2 | 79 | 1 | 340 | 1 |
| 140 | 1 | 80 | 1 | 360 | 1 |
| 170 | 1 | | --62 | 380 | 1 |
| 180 | 1 | | | 420 | 3 |
| 190 | 2 | | | 440 | 1 |
| 210 | 1 | | | 460 | 2 |
| 220 | 1 | | | 500 | 1 |
| 230 | 1 | | | 530 | 1 |
| 240 | 1 | | | 550 | 1 |
| 250 | 1 | | | 560 | 1 |
| 290 | 2 | | | 610 | 1 |
| 310 | 1 | | | 680 | 1 |
| 320 | 2 | | | 740 | 1 |
| 340 | 1 | | | 790 | 1 |
| 380 | 1 | | | 880 | 1 |
| 470 | 1 | | | 890 | 1 |
| 590 | 1 | | | 1000 | 1 |
| 760 | 1 | | | 1100 | 2 |
| 5200 | 1 | | | 1200 | 1 |
| | --92 | | | 1500 | 1 |
| | | | | 2700 | 2 |
| | | | | | --47 |

Tableau 52. Précision des tests d'ajustement de la distribution binomiale négative aux données sur les coliformes totaux de 1968, basés sur des estimations par MM et MV de p et k

| 1968, mission 102 | | | | | | 1968, mission 108 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Méthode des moments | | | Maximum de vraisemblance | | | Méthode des moments | | | Maximum de vraisemblance | | |
| r | f | e | r | f | e | r | f | e | r | f | e |
| 0 | 29 | 34.07 | 0 | 29 | 30.49 | 0 | 13 | 20.29 | 0 | 13 | 12.28 |
| 1 | 9 | 5.33 | 1 | 9 | 5.98 | 1 | 4 | 5.27 | 1,2 | 7 | 8.79 |
| 2,3 | 7 | 5.14 | 2-4 | 9 | 7.89 | 2-4 | 6 | 7.69 | 3,4 | 3 | 5.26 |
| 4-76 | 7 | 5.04 | 5-8 | 5 | 5.02 | 5-8 | 8 | 5.46 | 5-7 | 7 | 5.56 |
| 8-16 | 5 | 5.14 | 9-17 | 5 | 5.47 | 9-14 | 9 | 5.02 | 8-12 | 8 | 6.37 |
| 17-44 | 2 | 5.05 | 18-43 | 2 | 5.09 | 15-24 | 9 | 5.08 | 13-19 | 8 | 5.94 |
| 45-61 | 1 | 1.02 | 44-60 | 1 | 1.04 | 25-43 | 5 | 5.25 | 20-29 | 4 | 5.47 |
| 62 | 1 | 1.24 | 61 | 2 | 1.01 | 44-90 | 7 | 5.11 | 30-45 | 4 | 5.09 |
| | | | | | | 91-113 | 0 | 1.03 | 46-87 | 7 | 5.18 |
| | | | | | | 114 | 1 | 1.80 | 88-112 | 0 | 1.04 |
| | | | | | | | | | 113 | 1 | 1.02 |

| 1968, mission 109 | | | | | | 1968, mission 112 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Méthode des moments | | | Maximum de vraisemblance | | | Méthode des moments | | | Maximum de vraisemblance | | |
| r | f | e | r | f | e | r | f | e | r | f | e |
| 0 | 10 | 13.40 | 0 | 10 | 10.66 | 0 | 26 | 30.88 | 0 | 26 | 26.32 |
| 1-5 | 11 | 6.94 | 1,2 | 5 | 4.53 | 1-4 | 12 | 8.53 | 1,2 | 6 | 7.31 |
| 6-33 | 5 | 5.40 | 3-9 | 6 | 5.50 | 5-13 | 5 | 5.09 | 3-6 | 9 | 5.60 |
| 34-52 | 1 | 1.03 | 10-35 | 5 | 5.27 | 14-47 | 7 | 5.08 | 7-16 | 5 | 5.64 |
| ≥53 | 1 | 1.24 | 36-52 | 1 | 1.01 | 46-69 | 1 | 1.05 | 17-47 | 4 | 5.06 |
| | | | ≥53 | 1 | 1.03 | ≥70 | 1 | 1.37 | 48-68 | 1 | 1.04 |
| | | | | | | | | | ≥69 | 1 | 1.03 |

Précision du test d'ajustement

| | 68-102 | | 68-108 | | 68-109 | | 68-112 | |
|---|---|---|---|---|---|---|---|---|
| | MM | MV | MM | MV | MM | MV | MM | MV |
| Précision de la statistique d'ajustement | 7.04 | 4.63 | 12.77 | 5.19 | 3.32 | 0.15 | 3.01 | 2.61 |
| Degrés de liberté | 5 | 5 | 7 | 8 | 2 | 3 | 3 | 4 |
| Niveau de signification | 0.22 | 0.47 | 0.08 | 0.74 | 0.19 | 0.99 | 0.40 | 0.63 |
| Estimation de p | 41.10 | 32.58 | 68.07 | 41.59 | 40.90 | 34.61 | 62.67 | 42.62 |
| Estimation de k | 0.16 | 0.20 | 0.26 | 0.43 | 0.19 | 0.27 | 0.13 | 0.18 |

r - concentration bactérienne.
f - fréquence observée.
e - fréquence prévue.
MM- estimations par la méthode des moments.
MV- estimation par maximum de vraisemblance.

Tableau 53. Comparaison des méthodes de regroupement des données sur les coliformes totaux

| k | Moyenne du groupe | | | | | | | n | | | | | | | Indice de dispersion ($D^2$) | | | | | | | (a) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | -2 log A |
| **Méthode de regroupement par probabilités** | | | | | | | | | | | | | | | | | | | | | | |
| 2 | 2.09 | 57.8 | | | | | | 57 | 5 | | | | | | 332 (α.001) | 33.5 (<.001) | | | | | | 982.601 |
| 3 | 0.44 | 7.14 | 57.8 | | | | | 43 | 14 | 5 | | | | | 46.6 (>.30) | 27.7 (.01) | 33.5 (.001) | | | | | 128.014 |
| 4 | 0.44 | 7.14 | 35.5 | 72.7 | | | | 43 | 14 | 2 | 3 | | | | 46.6 (>.30) | 27.7 (.01) | .01 (>.90) | 3.86 (>.20) | | | | 85.406 |
| 5 | 0.44 | 7.14 | 35.5 | 59.0 | 79.5 | | | 43 | 14 | 2 | 1 | 2 | | | 46.6 (>.30) | 27.7 (.01) | .01 (>.90) | 0 - | 0.60 (> 70) | | | 8.113 |
| 6 | 0.44 | 7.14 | 35.0 | 36.0 | 59.0 | 79.5 | | 43 | 14 | 1 | 1 | 1 | 2 | | 46.6 (>.30) | 26.7 (.01) | 0 - | 0 - | 0 - | 0.01 (>.90) | | 0.014 |
| 7 | 0.24 | 3.46 | 10.83 | 35.0 | 36.0 | 59.0 | 79.5 | 38 | 13 | 6 | 1 | 1 | 1 | 2 | 29.0 (>.70) | 7.29 (>.80) | 4.33 (.50) | 0 - | 0 - | 0 - | 0.01 (>.90) | |
| **Indice de dispersion**(b) | | | | | | | | | | | | | | | | | | | | | | |
| 4 | 0.57 | 7.83 | 35.50 | 72.67 | | | | 45 | 12 | 2 | 3 | | | | 59.6 (.06) | 21.96 (.04) | 0.01 (>.90) | 3.86 (>.10) | | | | |

(a) El-Shaarawi, 1982, Tableau 2.
(b) El-Shaarawi et coll., 1981, Tableau 4.

Tableau 54a. Comparaison de l'ajustement de la distribution binomiale négative à certaines des données sur les coliformes totaux de 1967 effectuée à l'aide de toutes les données de la mission et d'un sous-ensemble dont les valeurs élevées ont été éliminées

| Mission 101 (n=150) | | | | Mission 103 (n=166) | | | | Mission 105 (n=170) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| r | f | e | $x^2$,(a) | r | f | e | $x^2$,(a) | r | f | e | $x^2$,(a) |
| 0 | 119 | 120.03 | 0.01 | 0 | 65 | 74.96 | 1.32 | 0 | 77 | 89.05 | 1.63 |
| 1-3 | 14 | 8.01 | 4.48 | 1 | 21 | 9.40 | 14.32 | 1 | 20 | 8.55 | 15.32 |
| 4-12 | 7 | 5.86 | 0.22 | 2 | 5 | 5.28 | 0.02 | 2,3 | 26 | 7.95 | 40.98 |
| 13-16 | 2 | 1.30 | 0.38 | 3,4 | 12 | 6.65 | 4.30 | 4-6 | 17 | 6.35 | 17.88 |
| ≥17 | 8 | 1.02 | 47.94 | 5-7 | 10 | 6.24 | 2.27 | 7-10 | 6 | 5.19 | 0.13 |
| | | | 53.02 | 8-11 | 11 | 5.53 | 5.42 | 11-16 | 10 | 5.10 | 4.70 |
| | | | (<<0.001) | 12-17 | 8 | 5.69 | 0.94 | 17,18 | 1 | 1.32 | 0.08 |
| | | | | ≥18 | 34 | 2.21 | 457.88 | >>19 | 13 | 1.20 | 116.50 |
| | | | | | | | 486.45 | | | | 197.21 |
| | | | | | | | (<<0.001) | | | | (<<0.001) |

Tableau 54b. Comparaison de l'ajustement de la distribution binomiale à certaines des données sur les coliformes totaux de 1967 effectuée à l'aide de toutes les données de la mission et d'un sous-ensemble dont les valeurs élevées ont été éliminées

| Mission 101 (n=146) | | | | Mission 103 (n=161) | | | | Mission 105 (n=165) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| r | f | e | $x^2$,(a) | r | f | e | $x^2$,(a) | r | f | e | $x^2$,(a) |
| 0 | 119 | 119.62 | 0.00 | 0 | 65 | 65.94 | 0.01 | 0 | 77 | 78.39 | 0.02 |
| 1 | 13 | 6.09 | 7.84 | 1 | 21 | 15.21 | 2.20 | 1 | 20 | 19.93 | 0.00 |
| 2-4 | 4 | 6.79 | 1.15 | 2 | 5 | 9.18 | 1.90 | 2 | 17 | 11.86 | 2.23 |
| 5-12 | 4 | 6.08 | 0.71 | 3 | 6 | 6.69 | 0.07 | 3 | 9 | 8.40 | 0.04 |
| 13-41 | 3 | 5.38 | 1.05 | 4,5 | 8 | 9.67 | 0.29 | 4 | 8 | 6.42 | 0.39 |
| 42-61 | 1 | 1.02 | 0.00 | 6,7 | 8 | 6.98 | 0.15 | 5 | 1 | 5.13 | 3.32 |
| ≥62 | 2 | 1.02 | 0.94 | 8,9 | 8 | 5.44 | 1.20 | 6,7 | 11 | 7.75 | 1.36 |
| | | | 11.72 | 10-12 | 8 | 6.36 | 0.42 | 8-10 | 3 | 7.83 | 2.98 |
| | | | (0.02) | 13-16 | 3 | 6.35 | 1.77 | 11-14 | 5 | 6.54 | 0.36 |
| | | | | 17-21 | 1 | 5.83 | 4.00 | 15-20 | 8 | 5.55 | 1.08 |
| | | | | 22-28 | 9 | 5.79 | 1.78 | 21-34 | 3 | 5.07 | 0.85 |
| | | | | 29-37 | 8 | 5.01 | 1.78 | 35-43 | 1 | 1.11 | 0.01 |
| | | | | 38-53 | 5 | 5.29 | 0.02 | ≥44 | 2 | 1.02 | 0.94 |
| | | | | 54-94 | 4 | 5.19 | 0.27 | | | | 13.58 |
| | | | | 95-119 | 1 | 1.05 | 0.00 | | | | (0.19) |
| | | | | ≥120 | 1 | 1.00 | 0.00 | | | | |
| | | | | | | | 15.89 | | | | |
| | | | | | | | (0.26) | | | | |

r,f,e,$x^2$ - le nombre de coliformes par 100 mL, la fréquence observée, la fréquence prévue et la valeur de $(f-e)^2/e$. La somme de ces derniers termes et le niveau de signification correspondant, a, sont indiqués au bas de chaque colonne des $x^2$.

n - nombre des observations.

Note: On a utilisé les estimations par maximum de vraisemblance de p et k.

Tableau 55. Test d'ajustement de la distribution binomiale négative aux données sur les coliformes totaux de 1967, 1968 et 1969

| | | | Méthode des moments | | | Maximum de vraisemblance | | |
|---|---|---|---|---|---|---|---|---|
| Année | Mission | n | $x^2$ | Degrés de liberté | α | $x^2$ | Degrés de liberté | α |
| 1967 | 101 | 146 | 3.76 | 3 | 0.29 | 11.72 | 4 | 0.02 |
| | 103 | 161 | 15.56 | 13 | 0.28 | 15.89 | 13 | 0.26 |
| | 105 | 165 | 12.83 | 9 | 0.18 | 13.58 | 10 | 0.19 |
| | 107 | 98 | 28.90 | 5 | <0.001 | 11.08 | 6 | 0.09 |
| | 109 | 95 | 100.01 | 4 | <0.001 | 25.01 | 7 | <0.001 |
| | 111 | 94 | 4.91 | 4 | 0.30 | 1.90 | 4 | 0.75 |
| | 113 | 91 | 36.03 | 4 | <0.001 | 8.53 | 6 | 0.20 |
| | 115 | 73 | 7.14 | 5 | 0.21 | 6.78 | 5 | 0.24 |
| 1968 | 102 | 62 | 7.04 | 5 | 0.22 | 4.63 | 5 | 0.47 |
| | 108 | 62 | 12.77 | 7 | 0.08 | 5.19 | 8 | 0.74 |
| | 109 | 28 | 3.32 | 2 | 0.19 | 0.15 | 3 | 0.99 |
| | 112 | 52 | 3.01 | 3 | 0.40 | 2.61 | 4 | 0.63 |
| 1969 | 103 | 54 | 9.25 | 5 | 0.10 | 2.50 | 5 | 0.77 |
| | 110 | 68 | 11.76 | 8 | 0.17 | 11.82 | 8 | 0.17 |

Tableau 56. Estimations des paramètres de la distribution binomiale négative pour les données sur les coliformes totaux de 1967, 1968 et 1969

| Année | Mission | n | Valeurs exclues | Méthode des moments | | Maximum de vraisemblance | | Moyenne | Variance | Médiane | Nombre de valeurs nulles |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $\tilde{p}$ | $\tilde{k}$ | $\hat{p}$ | $\hat{k}$ | | | | |
| 1967 | 101 | 146 | 230,520,610,1000 | 54.32 | 0.04 | 45.15 | 0.05 | 2.35 | 129.97 | 0 | 119 |
| | 103 | 161 | 270,330,530,600,8300 | 41.35 | 0.25 | 42.92 | 0.24 | 10.13 | 429.00 | 1 | 65 |
| | 105 | 165 | 300,340,500,800,11 000 | 17.06 | 0.23 | 14.44 | 0.27 | 3.93 | 70.92 | 1 | 77 |
| | 107 | 98 | 6300 | 46.87 | 0.10 | 25.20 | 0.19 | 4.86 | 232.52 | 0 | 50 |
| | 109 | 95 | 750,22 000 | 93.35 | 0.07 | 29.36 | 0.21 | 6.19 | 583.98 | 1 | 41 |
| | 111 | 94 | 900,1000,150 000 | 10.10 | 0.18 | 7.90 | 0.22 | 1.78 | 19.72 | 0 | 57 |
| | 113 | 91 | 500,6000,12 000, 51 000 | 74.97 | 0.07 | 31.05 | 0.18 | 5.54 | 420.74 | 0 | 47 |
| | 115 | 73 | 500,2600 | 29.41 | 0.21 | 30.46 | 0.20 | 6.07 | 184.54 | 1 | 36 |
| 1968 | 102 | 62 | | 41.10 | 0.16 | 32.58 | 0.20 | 6.58 | 277.03 | 1 | 29 |
| | 108 | 62 | | 68.07 | 0.26 | 41.59 | 0.43 | 17.95 | 1239.92 | 8.5 | 13 |
| | 109 | 28 | | 49.90 | 0.19 | 34.61 | 0.27 | 9.36 | 476.31 | 2 | 10 |
| | 112 | 52 | | 62.67 | 0.13 | 43.62 | 0.18 | 7.87 | 500.82 | 0.5 | 26 |
| 1969 | 103* | 54 | 360,700,4200,4900,7000 | 60.34 | 0.21 | 76.37 | 0.17 | 12.74 | 781.48 | 0.5 | 27 |
| | 110* | 68 | 210,480,520,540,770 | 41.00 | 0.36 | 43.40 | 0.34 | 14.66 | 615.78 | 3 | 18 |

* Les valeurs indéterminées ont été converties en zéros.

Tableau 57. Test d'ajustement de la distribution lognormale aux données sur les coliformes totaux de la mission 111 de 1966 et de la mission 107 de 1967

| Mission 111, 1966 | | | | Mission 107, 1970 | | | |
|---|---|---|---|---|---|---|---|
| Limites des classes $\log_e$ R | Fréquence observée f | Fréquence prévue e | $\frac{(f-e)^2}{e}$ | Limites des classes $\log_e$ R | Fréquence observée f | Fréquence prévue e | $\frac{(f-e)^2}{e}$ |
| <0.5 | 23 | 14.27 | 5.33 | ≤4.5 | 5 | 4.05 | 0.23 |
| [0.5,1.0) | 4 | 5.62 | 0.47 | [4.5,5.0) | 5 | 5.11 | 0.00 |
| [1.0,1.5) | 4 | 6.73 | 1.11 | [5.0,5.5) | 6 | 7.82 | 0.42 |
| [1.5,2.0) | 4 | 7.64 | 1.73 | [5.5,6.0) | 8 | 9.32 | 0.19 |
| [2.0,2.5) | 15 | 8.23 | 5.57 | [6.0,6.5) | 11 | 8.66 | 0.64 |
| [2.5,3.0) | 3 | 8.41 | 3.48 | [6.5,7.0) | 6 | 6.26 | 0.01 |
| [3.0,3.5) | 8 | 8.16 | 0.00 | ≥7.0 | 6 | 5.79 | 0.01 |
| [3.5,4.0) | 5 | 7.51 | 0.84 | | | $\chi^2_4$ = | 1.49 |
| [4.0,4.5) | 2 | 6.55 | 3.16 | | | | |
| [4.5,5.0) | 4 | 5.43 | 0.38 | | | $\alpha$ = | 0.83 |
| [5.0,5.5) | 8 | 4.27 | 3.27 | | | | |
| [5.5,6.5) | 10 | 5.43 | 3.85 | | | | |
| ≥6.5 | 2 | 3.75 | 0.82 | | | | |
| | | $\chi^2_{10}$ = | 30.00 | | | | |
| | | $\alpha$ = | 0.001 | | | | |

[a,b) indique que la concentration est ≥a et <b.

Tableau 58. Zones 1 à 5 des coliformes totaux par mission, de 1966 à 1970

| Mission | Date | Regroupement en zones par rapport de vraisemblance: Moyenne de la zone 1 | 2 | 3 | 4 | Nombre dans la zone 1 | 2 | 3 | 4 | Zone 5 | Zones 1 à 4: Min. | Max. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 66111 | 8-14 août | 5.20 | 56.33 | 251.18 | 606.67 | 56 | 15 | 17 | 3 | 5200 | 0 | 760 |
| 67101 | 30 mai - 8 juin | 0.10 | 7.50 | 40.00 | 87.50 | 132 | 10 | 2 | 2 | 230,520,610,1000 | 0 | 100 |
| 67103 | 20-29 juin | 1.16 | 19.68 | 61.20 | 160.00 | 113 | 37 | 10 | 1 | 270,330,530,600,8300 | 0 | 160 |
| 67105 | 10-19 juil. | 0.21 | 3.81 | 14.86 | 39.50 | 97 | 48 | 14 | 6 | 300,340,500,800,11 000 | 0 | 60 |
| 67107 | 31 juil. - 10 août | 0.36 | 6.44 | 34.33 | 100.00 | 70 | 23 | 3 | 2 | 6300 | 0 | 100 |
| 67109 | 21-30 août | 0.52 | 6.41 | 37.00 | 220.00 | 71 | 17 | 6 | 1 | 750,22 000 | 0 | 220 |
| 67111 | 11-17 sept. | 0.33 | 6.29 | 35.00 | - | 76 | 17 | 1 | - | 900,1000,150 000 | 0 | 35 |
| 67113 | 2-9 oct. | 0.18 | 4.22 | 33.33 | 180.00 | 57 | 27 | 6 | 1 | 500,6000,12 000,51 000 | 0 | 180 |
| 67115 | 24-30 oct. | 0.33 | 6.50 | 40.00 | 70.00 | 48 | 18 | 6 | 1 | 500,2600 | 0 | 70 |
| 68102 | 17-22 mai | 0.44 | 7.14 | 43.33 | 79.50 | 43 | 14 | 3 | 2 | | 0 | 80 |
| 68108 | 29 juil. - 3 août | 1.52 | 13.23 | 43.42 | 260.00 | 27 | 22 | 12 | 1 | | 0 | 260 |
| 68109 | 31 août - 2 sept. | 1.33 | 14.25 | 33.50 | 110.00 | 24 | 4 | 2 | 1 | | 0 | 110 |
| 68112 | 5-10 nov. | 0.98 | 14.67 | 44.00 | 150.00 | 40 | 9 | 2 | 1 | | 0 | 150 |
| 69103* | 30 mai - 4 juin | 0.19 | 10.31 | 51.40 | 130.00 | 31 | 16 | 5 | 2 | 360,700,4200,4900,7000 | 0 | 140 |
| 69110* | 15-20 oct. | 2.02 | 23.50 | 63.57 | 130.00 | 46 | 14 | 7 | 1 | 210,480,520,540,770 | 0 | 130 |
| 70107 | 28 juil. - 2 août | 144.35 | 439.47 | 1022.22 | 2700.00 | 17 | 19 | 9 | 2 | | 20 | 2700 |

* Valeurs indéterminées transformées en 0.

Tableau 59. Concentrations de coliformes totaux et emplacements des stations de la zone 5

| Année et mission | Concentration dans les environs de | | | | Concentration | Emplacement |
|---|---|---|---|---|---|---|
| | Toledo | Cleveland | Erie | Rivière Grand Ontario | | |
| 1966.111 | | | | | 5200 | Centre du bassin est |
| 1967.101 | | 230,520,610,1000 | | | | |
| 103 | | 270,330,530,8300 | | 600 | | |
| 105 | 800 | 11 000 | 500 | 300 | 340 | Extrémité S.E. du bassin est |
| 107 | | 6300 | | | | |
| 109 | 750 | 22 000 | | | | |
| 111 | 150 000 | 900 | | 1000 | | |
| 113 | 51 000 | 12 000 | 500 | 6000 | | |
| 115 | | 2600 | | 500 | | |
| 1969.103 | | 4200,4900,7000 | | | 360 | Long Point |
| | | | | | 700 | Rivière Détroit |
| 110 | | 210,520 | | | 480 | Pointe-Pelée |
| | | | | | 540,770 | Rivière Détroit |

Tableau 60. Distributions des fréquences observées des données sur les coliformes féaux, 1967 à 1975

| r | Fréquence | f(r) |
|---|---|---|
| **67-103** | | |
| 0 | 34 | 0.723 |
| 1 | 1 | 0.021 |
| 2 | 2 | 0.043 |
| 4 | 2 | 0.043 |
| 6 | 1 | 0.021 |
| 8 | 1 | 0.021 |
| 28 | 1 | 0.021 |
| 34 | 1 | 0.021 |
| 68 | 1 | 0.021 |
| 92 | 1 | 0.021 |
| 120 | 1 | 0.021 |
| 400 | 1 | 0.021 |
| | n = 47 | 1.000 |
| **67-105** | | |
| 0 | 42 | 0.737 |
| 2 | 8 | 0.140 |
| 8 | 1 | 0.018 |
| 10 | 2 | 0.035 |
| 16 | 1 | 0.018 |
| 90 | 1 | 0.018 |
| 300E | 1 | 0.018 |
| 1600C | 1 | 0.018 |
| | 57 | 1.000 |
| **67-107** | | |
| 0 | 26 | 0.839 |
| 2 | 2 | 0.065 |
| 14 | 1 | 0.032 |
| 44 | 1 | 0.032 |
| 5000C | 1 | 0.032 |
| | 31 | 1.000 |
| **67-109** | | |
| 0 | 26 | 0.867 |
| 4 | 1 | 0.033 |
| 32 | 1 | 0.033 |
| 250C | 1 | 0.033 |
| 250W | 1 | 0.033 |
| | 30 | 1.000 |
| **67-111** | | |
| 0 | 22 | 0.815 |
| 1 | 1 | 0.037 |
| 2 | 1 | 0.037 |
| 4 | 1 | 0.037 |
| 40 | 1 | 0.037 |
| 400G | 1 | 0.037 |
| | 27 | 1.000 |
| **67-113** | | |
| 0 | 17 | 0.515 |
| 2 | 4 | 0.121 |
| 4 | 1 | 0.030 |
| 6 | 1 | 0.030 |
| 8 | 1 | 0.030 |
| 30 | 1 | 0.030 |
| 34 | 1 | 0.030 |
| 150 | 1 | 0.030 |
| 190 | 1 | 0.030 |
| 880G | 1 | 0.030 |
| 3 500C | 1 | 0.030 |
| 12 000T | 1 | 0.030 |
| | 33 | 1.000 |

| r | Fréquence | f(r) |
|---|---|---|
| **67-115** | | |
| 0 | 14 | 0.667 |
| 1 | 1 | 0.048 |
| 2 | 2 | 0.095 |
| 6 | 1 | 0.048 |
| 17 | 1 | 0.048 |
| 42 | 1 | 0.048 |
| 340C | 1 | 0.048 |
| | 21 | 1.000 |
| **68-102** | | |
| 0 | 22 | 0.629 |
| 1 | 4 | 0.114 |
| 2 | 2 | 0.057 |
| 3 | 2 | 0.057 |
| 6 | 1 | 0.029 |
| 18 | 1 | 0.029 |
| 19 | 1 | 0.029 |
| 53 | 1 | 0.029 |
| 54 | 1 | 0.029 |
| | 35 | 1.000 |
| **68-108** | | |
| 0 | 27 | 0.771 |
| 2 | 2 | 0.057 |
| 4 | 1 | 0.029 |
| 6 | 1 | 0.029 |
| 7 | 1 | 0.029 |
| 15 | 2 | 0.057 |
| 24 | 1 | 0.029 |
| | 35 | 1.000 |
| **68-112** | | |
| 0 | 16 | 0.533 |
| 2 | 4 | 0.133 |
| 4 | 2 | 0.067 |
| 5 | 1 | 0.033 |
| 6 | 1 | 0.033 |
| 9 | 1 | 0.033 |
| 13 | 1 | 0.033 |
| 14 | 1 | 0.033 |
| 72 | 1 | 0.033 |
| 90 | 1 | 0.033 |
| 350C | 1 | 0.033 |
| | 30 | 1.000 |
| **69-110** | | |
| 0 | 58 | 0.806 |
| 1 | 6 | 0.083 |
| 2 | 2 | 0.028 |
| 3 | 1 | 0.014 |
| 4 | 1 | 0.014 |
| 6 | 1 | 0.014 |
| 8 | 1 | 0.014 |
| 10 | 1 | 0.014 |
| 48 | 1 | 0.014 |
| | 72 | 1.000 |

| r | Fréquence | f(r) |
|---|---|---|
| **74-104** | | |
| 0 | 20 | 0.667 |
| 1 | 1 | 0.033 |
| 2 | 1 | 0.033 |
| 3 | 1 | 0.033 |
| 6 | 1 | 0.033 |
| 9 | 1 | 0.033 |
| 10 | 1 | 0.033 |
| 13 | 1 | 0.033 |
| 21 | 1 | 0.033 |
| 34 | 1 | 0.033 |
| 53 | 1 | 0.033 |
| | 30 | 1.000 |
| **75-101** | | |
| 0 | 25 | 0.781 |
| 1 | 4 | 0.125 |
| 27 | 1 | 0.031 |
| 39 | 1 | 0.031 |
| 210L | 1 | 0.031 |
| | 32 | 1.000 |
| **75-106** | | |
| 0 | 8 | 0.667 |
| 1 | 1 | 0.083 |
| 3 | 2 | 0.167 |
| 16 | 1 | 0.083 |
| | 12 | 1.000 |
| **75-107** | | |
| 0 | 16 | 0.516 |
| 1 | 1 | 0.032 |
| 2 | 1 | 0.032 |
| 5 | 1 | 0.032 |
| 7 | 1 | 0.032 |
| 10 | 1 | 0.032 |
| 43 | 1 | 0.032 |
| 76 | 1 | 0.032 |
| 83 | 1 | 0.032 |
| 100 | 1 | 0.032 |
| 110 | 2 | 0.065 |
| 180 | 1 | 0.032 |
| 190 | 1 | 0.032 |
| 310E | 1 | 0.032 |
| 430C | 1 | 0.032 |
| | 31 | 1.000 |

Note: r = concentration de coliformes fécaux. Les lettres qui suivent la concentration indiquent l'emplacement de la station. Les lettres C, E, G, L et T représentent respectivement les environs de Cleveland, d'Erie, de la rivière Grand, de Long Point et de Toledo.

Tableau 61. Distribution des fréquences conditionnelles $f(R_f/R_t)$ des concentrations de coliformes fécaux, $R_f$, pour des concentrations données de coliformes totaux, $R_t$, données de 1967 et 1968

| | | Nombre | | | | Nombre | | | | | Nombre | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $R_t$ | $R_f$ | $R_t$ AR$_f$ | $R_t$ | $R_t$ | $R_f$ | $R_t$ AR$_f$ | $R_t$ | $R_t$ | $R_f$ | $R_t$ AR$_f$ | $R_t$ |
| 0 | 0 | 112 | 118 | 9 | 0 | 3 | 4 | [40,50) | 0 | 6 | 12 |
| | 1 | 1 | | | 4 | 1 | | | 2 | 1 | |
| | 2 | 4 | | | | | | | 4 | 1 | |
| | 40 | 1 | | 10 | 0 | 8 | 8 | | 16 | 2 | |
| | | | | | | | | | 15 | 1 | |
| 1 | 0 | 25 | 29 | 11 | 0 | 1 | 2 | | 42 | 1 | |
| | 2 | 4 | | | 1 | 1 | | | | | |
| | | | | | | | | [50,100) | [0,10) | 4 | 9 |
| 2 | 0 | 17 | 21 | 12 | 0 | 6 | 8 | | [10,20) | 1 | |
| | 1 | 2 | | | 2 | 1 | | | [20,30) | 2 | |
| | 2 | 2 | | | 8 | 1 | | | [30,40) | 0 | |
| | | | | | | | | | [40,50) | 0 | |
| 3 | 0 | 8 | 11 | 13 | 0 | 1 | 1 | | [50,60) | 2 | |
| | 2 | 1 | | | | | | | | | |
| | 6 | 1 | | 14 | 2 | 1 | 2 | [100,500) | [0,25) | 1 | 9 |
| | 10 | 1 | | | 6 | 1 | | | [25,50) | 4 | |
| | | | | | | | | | [50,100) | 3 | |
| 4 | 0 | 12 | 15 | 15 | 7 | 1 | 1 | | [100,150) | 1 | |
| | 2 | 2 | | | | | | | | | |
| | 3 | 1 | | 16 | 0 | 4 | 7 | [500,1000) | [0,50) | 2 | 6 |
| | | | | | 2 | 1 | | | [50,100) | 1 | |
| 5 | 0 | 10 | 13 | | 13 | 1 | | | [100,200) | 1 | |
| | 2 | 2 | | | 14 | 1 | | | [200,300) | 1 | |
| | 3 | 1 | | | | | | | [300,400) | 1 | |
| | | | | 17 | 9 | 1 | 1 | | | | |
| 6 | 0 | 5 | 8 | | | | | [1000,5000) | 340 | 1 | 2 |
| | 1 | 1 | | 18 | 2 | 1 | 2 | | 400 | 1 | |
| | 4 | 1 | | | 8 | 1 | | | | | |
| | 5 | 1 | | | | | | [5000,10 000) | 400 | 1 | 3 |
| | | | | [20,30) | 0 | 8 | 13 | | 880 | 1 | |
| 7 | 0 | 1 | 4 | | 2 | 3 | | | 5 000 | 1 | |
| | 1 | 1 | | | 4 | 1 | | | | | |
| | 2 | 1 | | | 52 | 1 | | [10 000,25 000) | 250 | 1 | 3 |
| | 8 | 1 | | | | | | | 1 600 | 1 | |
| | | | | [30,40) | 0 | 4 | 12 | | 3 500 | 1 | |
| 8 | 0 | 7 | 8 | | 2 | 2 | | | | | |
| | 4 | 1 | | | 4 | 2 | | 51 000 | 12 000 | 1 | 1 |
| | | | | | 6 | 1 | | | | | |
| | | | | | 10 | 1 | | | | | |
| | | | | | 14 | 1 | | | | | |
| | | | | | 18 | 1 | | | | | |

Les colonnes 3, 7 et 11 donnent le nombre de stations où, à la fois, la concentration de coliformes totaux est $R_t$, et celle de coliformes fécaux, $R_f$.
Les colonnes 4, 8 et 12 donnent le nombre de stations où la concentration de coliformes totaux est égale à $R_t$.
[a,b) indique que la concentration est $\geq$ a et < b.

Tableau 62. Résumé des concentrations de coliformes fécaux aux stations où les concentrations de coliformes totaux et fécaux ont été déterminées

| Mission | Date | Moyenne pour la zone 1 | 2 | 3 | 4 | Nombre dans la zone 1 | 2 | 3 | 4 | Zone 5 | Zones 1 à 4 Min. | Max. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 67103 | 20-29 juin | 0.17 | 1.23 | 7.00 | 34.00 | 24 | 13 | 5 | 1 | 68,92,120,400 | 0 | 34 |
| 67105 | 10-19 juil. | 0.36 | 1.00 | 1.43 | 2.50 | 28 | 14 | 7 | 4 | 16,90,300,1600 | 0 | 10 |
| 67107 | 31 juil.-10 août | 0.18 | 0.00 | 4.67 | 44.00 | 22 | 4 | 3 | 1 | 5000 | 0 | 44 |
| 67109 | 21-30 août | 0.00 | 0.00 | 1.33 | 32.00 | 16 | 8 | 3 | 1 | 250,260 | 0 | 32 |
| 67111 | 11-17 sept. | 2.563 (0.07)* | 0.50 | 2.00 | -- | 16 (15)* | 8 | 1 | 0 | 0,400 | 0 | 40 |
| 67113 | 2-9 oct. | 0.17 | 1.25 | 13.20 | 34.00 | 12 | 8 | 5 | 1 | 150,880,3500 12 000 | 0 | 52 |
| 67115 | 24-30 oct. | 0.18 | 1.40 | 14.67 | 17.00 | 11 | 5 | 3 | 1 | 340 | 0 | 42 |
| 68102 | 17-22 mai | 0.11 | 1.09 | 12.00 | 53.50 | 19 | 11 | 2 | 2 | | 0 | 54 |
| 68108 | 29 juil.-3 août | 0.00 | 1.25 | 6.43 | 15.00 | 13 | 12 | 7 | 1 | | 0 | 24 |
| 68112 | 5-10 nov. | 0.40 | 7.29 | -- | 72.00 | 15 | 7 | 0 | 1 | | 0 | 72 |
| 69110 | 15-20 oct. | 0.13 | 0.14 | 1.29 | 4.00 | 45 | 14 | 7 | 1 | 0,2,8,10,48 | 1 | 6 |

* Station où la concentration de coliformes fécaux était de 40 et qui a été éliminée.

Note: Les zones sont celles déterminées par l'inclusion de toutes les stations où la concentration de coliformes totaux a été déterminée. Voir tableau 58.

Tableau 63. Résumé des concentrations de coliformes totaux aux stations où les concentrations de coliformes totaux et fécaux ont été déterminées

| | Moyenne pour la zone | | | | Nombre dans la zone | | | | Zone | Zones 1 à 4 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Mission | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 5 | Min. | Max. |
| 67103 | 1.17 | 19.92 | 56.40 | 160.00 | 24 | 13 | 5 | 1 | 270,330,600,8300 | 0 | 160 |
| 67105 | 0.14 | 4.29 | 15.43 | 36.25 | 28 | 14 | 7 | 4 | 300,500,800,11 000 | 0 | 45 |
| 67107 | 0.50 | 9.50 | 34.33 | 100.00 | 22 | 4 | 3 | 1 | 6300 | 0 | 100 |
| 67109 | 0.44 | 6.00 | 36.33 | 220.00 | 16 | 8 | 3 | 1 | 750,22 000 | 0 | 220 |
| 67111 | 0.38 (0.40)* | 6.88 | 35.00 | -- | 16 (15)* | 8 | 1 | 0 | 900,1000 | 0 | 35 |
| 67113 | 0.17 | 4.50 | 28.40 | 180.00 | 12 | 8 | 5 | 1 | 500,6000,12 000 51 000 | 0 | 180 |
| 67115 | 0.27 | 4.00 | 38.00 | 70.00 | 11 | 5 | 3 | 1 | 2600 | 0 | 70 |
| 68102 | 0.63 | 7.46 | 35.50 | 79.50 | 19 | 11 | 2 | 2 | 79,80 | 0 | 80 |
| 68108 | 1.46 | 14.00 | 44.29 | 260.00 | 13 | 12 | 7 | 1 | | 0 | 260 |
| 68112 | 1.40 | 12.86 | -- | 150.00 | 15 | 7 | 0 | 1 | | 0 | 150 |
| 69110 | 2.04 | 23.50 | 63.57 | 130.00 | 45 | 14 | 7 | 1 | 210,480,520,540,770 | 1 | 130 |

* Station où la concentration de coliformes fécaux était de 40 et qui a été éliminée.

Note: Les zones sont celles déterminées par l'inclusion de toutes les stations où la concentration de coliformes totaux a été déterminée. Voir tableau 58.

Tableau 64. Rapport des concentrations de coliformes fécaux aux concentrations de coliformes totaux

| Mission | Zone | Toutes les stations | | Stations communes | | | Rapport | |
|---|---|---|---|---|---|---|---|---|
| | | Moyenne c. totaux (1) | n | Moyenne c. totaux (2) | Moyenne c. fécaux (3) | n | (3)/(1) | (3)/(2) |
| 67103 | 1 | 1.16 | 113 | 1.71 | 0.17 | 24 | 0.15 | 0.10 |
| | 2 | 19.68 | 37 | 19.92 | 1.23 | 13 | 0.06 | 0.06 |
| | 3 | 61.20 | 10 | 56.40 | 7.00 | 5 | 0.11 | 0.12 |
| 67105 | 1 | 0.21 | 97 | 0.14 | 0.36 | 28 | 1.71 | 2.57 |
| | 2 | 3.81 | 48 | 4.29 | 1.00 | 14 | 0.26 | 0.23 |
| | 3 | 14.86 | 14 | 15.43 | 1.43 | 7 | 0.10 | 0.09 |
| 67107 | 1 | 0.36 | 70 | 0.50 | 0.18 | 22 | 0.50 | 0.36 |
| | 2 | 6.44 | 23 | 9.50 | 0.00 | 4 | 0 | 0 |
| | 3 | 34.33 | 3 | 34.33 | 4.67 | 3 | 0.14 | 0.14 |
| 67109 | 1 | 0.52 | 71 | 0.44 | 0.00 | 16 | 0 | 0 |
| | 2 | 6.41 | 17 | 6.00 | 0.00 | 8 | 0 | 0 |
| | 3 | 37.00 | 6 | 36.33 | 1.33 | 3 | 0.04 | 0.04 |
| 67111 | 1 | 0.33 | 76 | 0.40 | 0.07 | 15 | 0.21 | 0.17* |
| | 2 | 6.29 | 17 | 6.88 | 0.50 | 8 | 0.08 | 0.07 |
| 67113 | 1 | 0.18 | 57 | 0.17 | 0.17 | 12 | 0.94 | 1.00 |
| | 2 | 4.22 | 27 | 4.50 | 1.25 | 8 | 0.30 | 0.28 |
| | 3 | 33.33 | 6 | 28.40 | 13.20 | 5 | 0.40 | 0.46 |
| 67115 | 1 | 0.33 | 48 | 0.27 | 0.18 | 11 | 0.55 | 0.67 |
| | 2 | 6.50 | 18 | 4.00 | 1.40 | 5 | 0.22 | 0.35 |
| | 3 | 40.00 | 6 | 38.00 | 14.67 | 3 | 0.37 | 0.39 |
| 68102 | 1 | 0.44 | 43 | 0.63 | 0.11 | 19 | 0.25 | 0.17 |
| | 2 | 7.14 | 14 | 7.46 | 1.09 | 11 | 0.15 | 0.15 |
| | 3 | 43.33 | 3 | 35.50 | 12.00 | 2 | 0.28 | 0.34 |
| 68108 | 1 | 1.52 | 27 | 1.46 | 0.00 | 13 | 0 | 0 |
| | 2 | 13.23 | 22 | 14.00 | 1.25 | 12 | 0.09 | 0.09 |
| | 3 | 43.42 | 12 | 44.29 | 6.43 | 7 | 0.15 | 0.15 |
| 68112 | 1 | 0.98 | 40 | 1.40 | 0.40 | 15 | 0.41 | 0.29 |
| | 2 | 14.67 | 9 | 12.86 | 7.29 | 7 | 0.50 | 0.57 |
| | 3 | 44.00 | 2 | - | - | 0 | - | - |
| 69110 | 1 | 3.02 | 46 | 3.07 | 1.02 | 45 | 0.34 | 0.33 |
| | 2 | 23.50 | 14 | 23.50 | 1.00 | 14 | 0.04 | 0.04 |
| | 3 | 63.57 | 7 | 63.57 | 2.00 | 7 | 0.03 | 0.03 |

* Station où la concentration de coliformes fécaux était de 40 et qui a été exclue.

Tableau 65. Turbidité et coliformes totaux par zones utilisant seulement les stations où la turbidité et les coliformes totaux ont été déterminés

| | | Moyenne pour la zone | | | | Nombre dans la zone | | | | Zones 1 à 4 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Variable | Mission | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | Min. | Max. |
| Turbidité | 68102 | 1.20 | 2.04 | 2.27 | 6.00 | 43 | 12 | 3 | 2 | 0.1 | 9.2 |
| | 68108 | 0.33 | 0.46 | 0.51 | 0.20 | 27 | 22 | 11 | 1 | 0.1 | 1.8 |
| | 68109 | 0.65 | 0.65 | - | - | 23 | 5 | 0 | 0 | 0.0 | 3.0 |
| | 68112 | 1.74 | 3.29 | 2.35 | 12.00 | 40 | 9 | 2 | 1 | 0.2 | 12.0 |
| Coliformes totaux | 68102 | 0.44 | 7.17 | 43.33 | 79.50 | 43 | 12 | 3 | 2 | 0 | 80 |
| | 68108 | 1.52 | 13.23 | 44.82 | 260.00 | 21 | 22 | 11 | 1 | 0 | 260 |
| | 68109 | 2.83 | 48.50 | - | - | 23 | 4 | 0 | 0 | 0 | 110 |
| | 68112 | 0.98 | 14.67 | 44.00 | 150.00 | 40 | 9 | 2 | 1 | 0 | 150 |

Les nombres totaux de stations communes sont 60, 61, 27 et 52 pour les missions 102, 108, 109 et 112, comparativement à 62, 62, 28, et 52 pour les coliformes totaux.

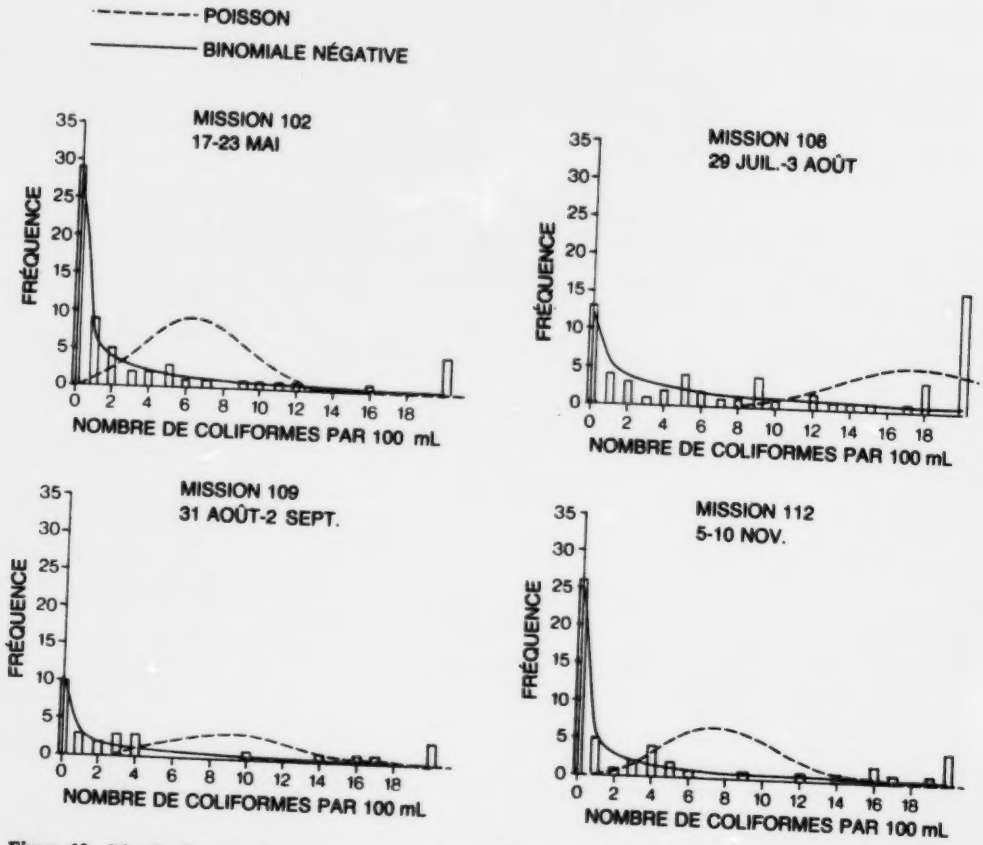



**Figure 63. Distribution des fréquences observées et distributions de Poisson et binomiale négative ajustées, données sur les coliformes totaux de 1968.**

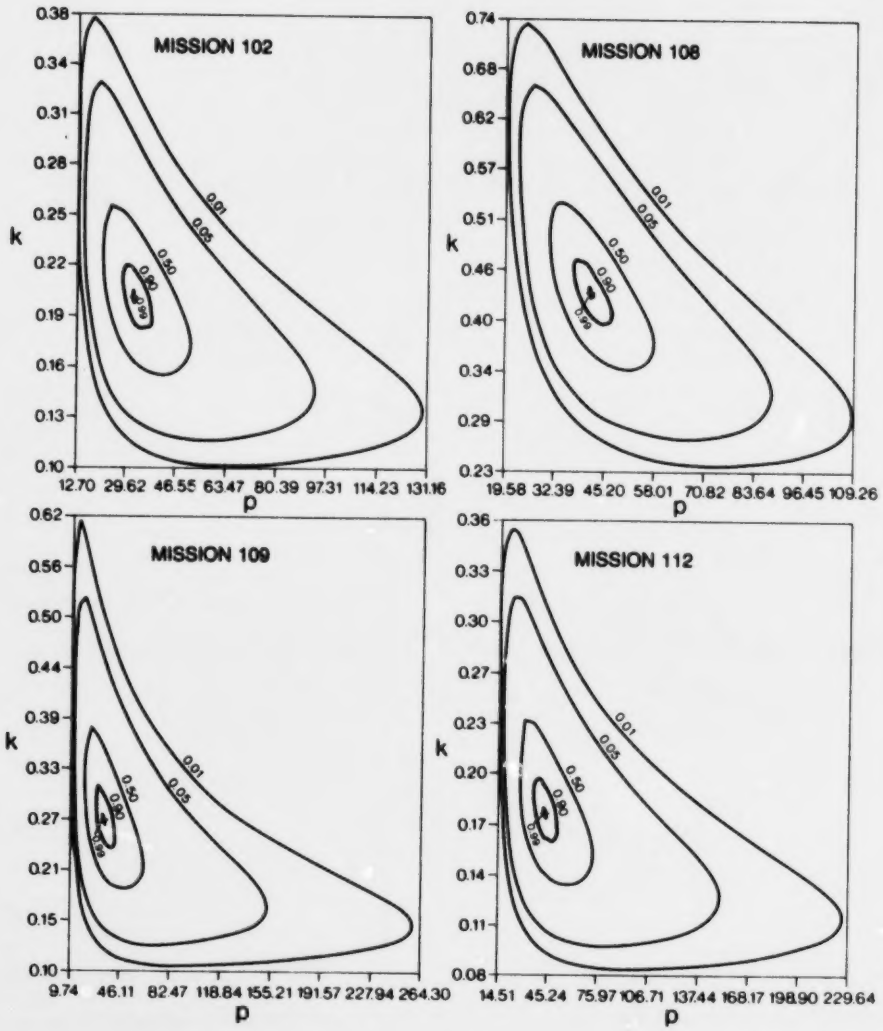



**Figure 64. Isoligne de vraisemblance relative R(p,k) = c, c = 0.01, 0.05, 0.50, 0.90, 0.99, pour les paramètres p et k de la distribution binomiale négative et les données sur les coliformes totaux — missions 102, 108, 109 et 112 de 1968.**

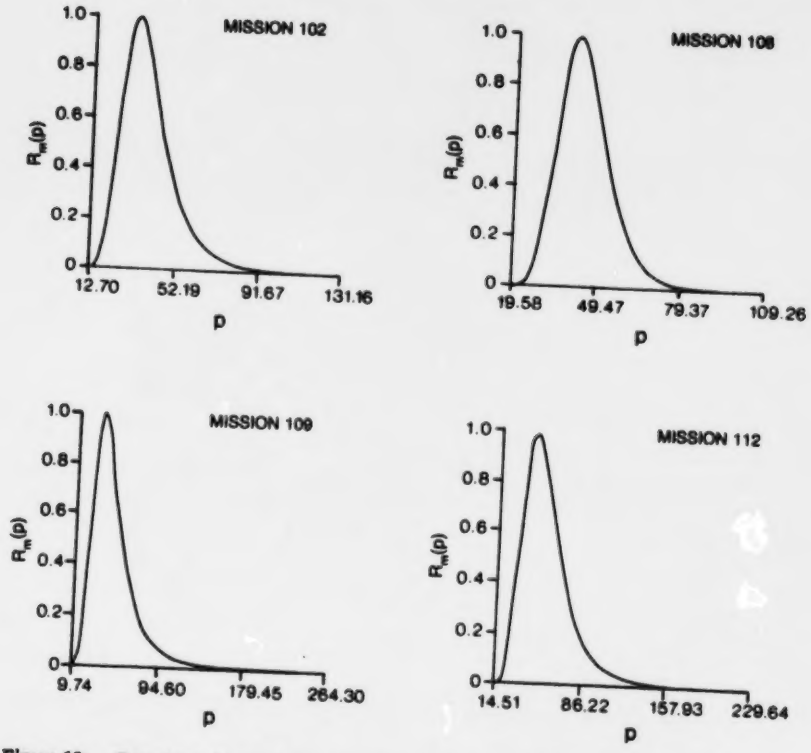



**Figure 65a.** Fonctions de maximum de vraisemblance relatif de $R_m(p)$ en fonction de p, un paramètre de la distribution binomiale négative, et la concentration de coliformes totaux — missions 102, 108, 109 et 112 de 1968.

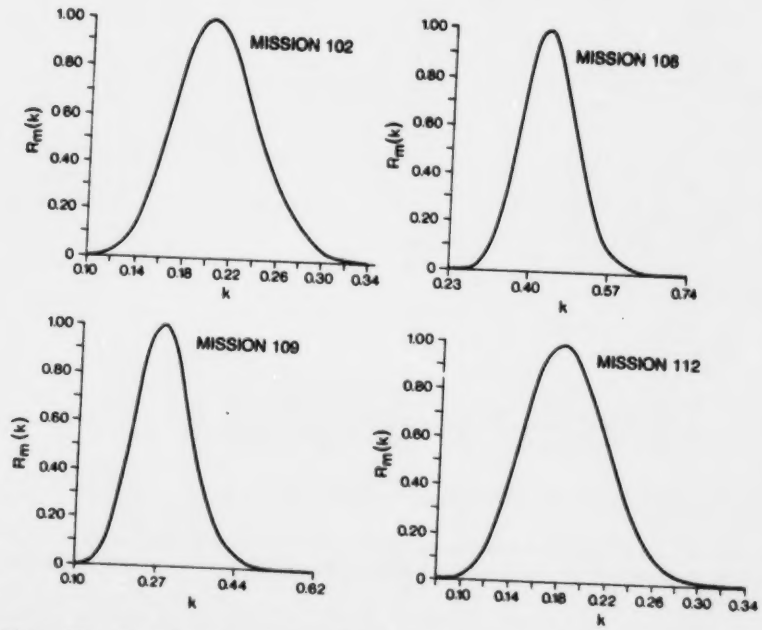
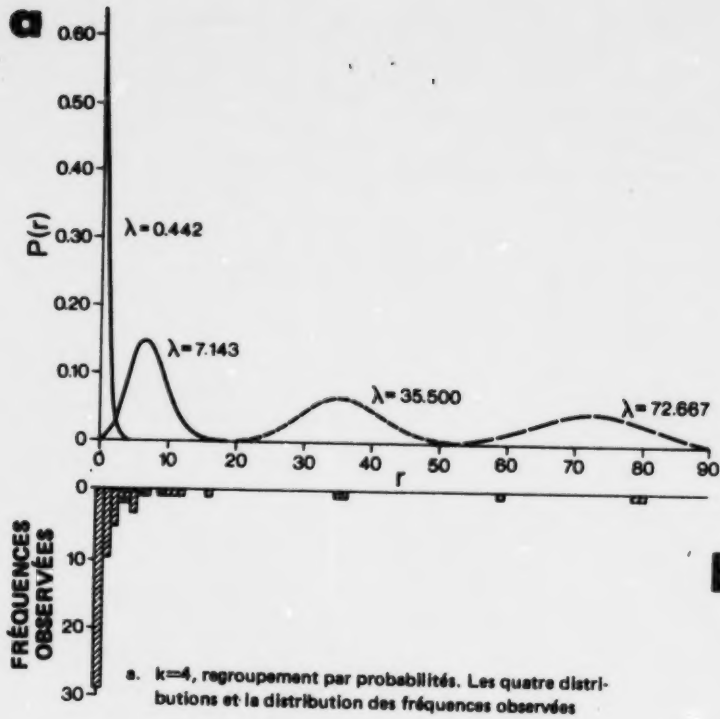
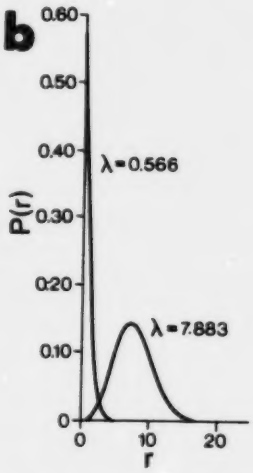


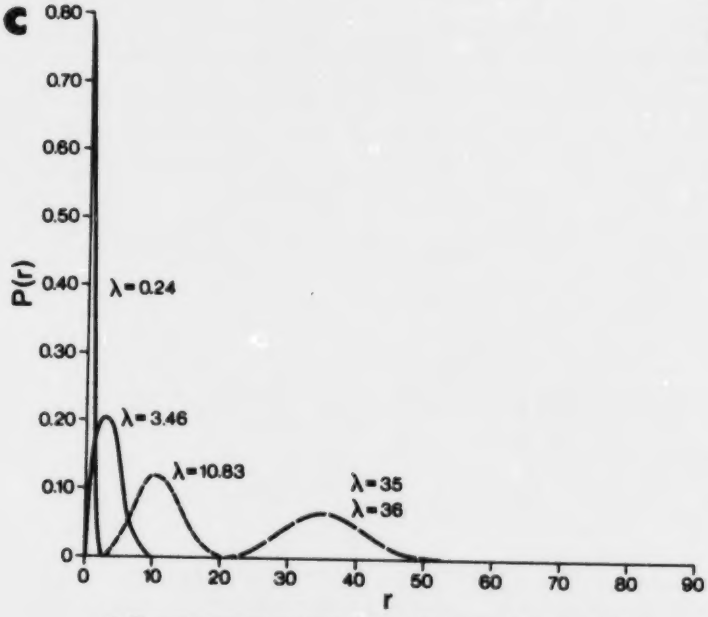

**Figure 65b.** Fonctions de maximum de vraisemblance relatif de $R_m(k)$ en fonction de k, un paramètre de la distribution binomiale negative, et la concentration de coliformes totaux — missions 102, 108, 109 et 112 de 1968.




a. k=4, regroupement par probabilités. Les quatre distributions et la distribution des fréquences observées




b. k=4, regroupement par le critère $D^3$, deux premières distributions




c. k=7, regroupement par probabilités. Cinq premières distributions

**Figure 66. Distributions des probabilités correspondant aux groupes déterminés pour la mission 102 (1968) où P (r) est la probabilité en supposant une distribution de Poisson de moyenne λ indiquée adjacente à la courbe.**

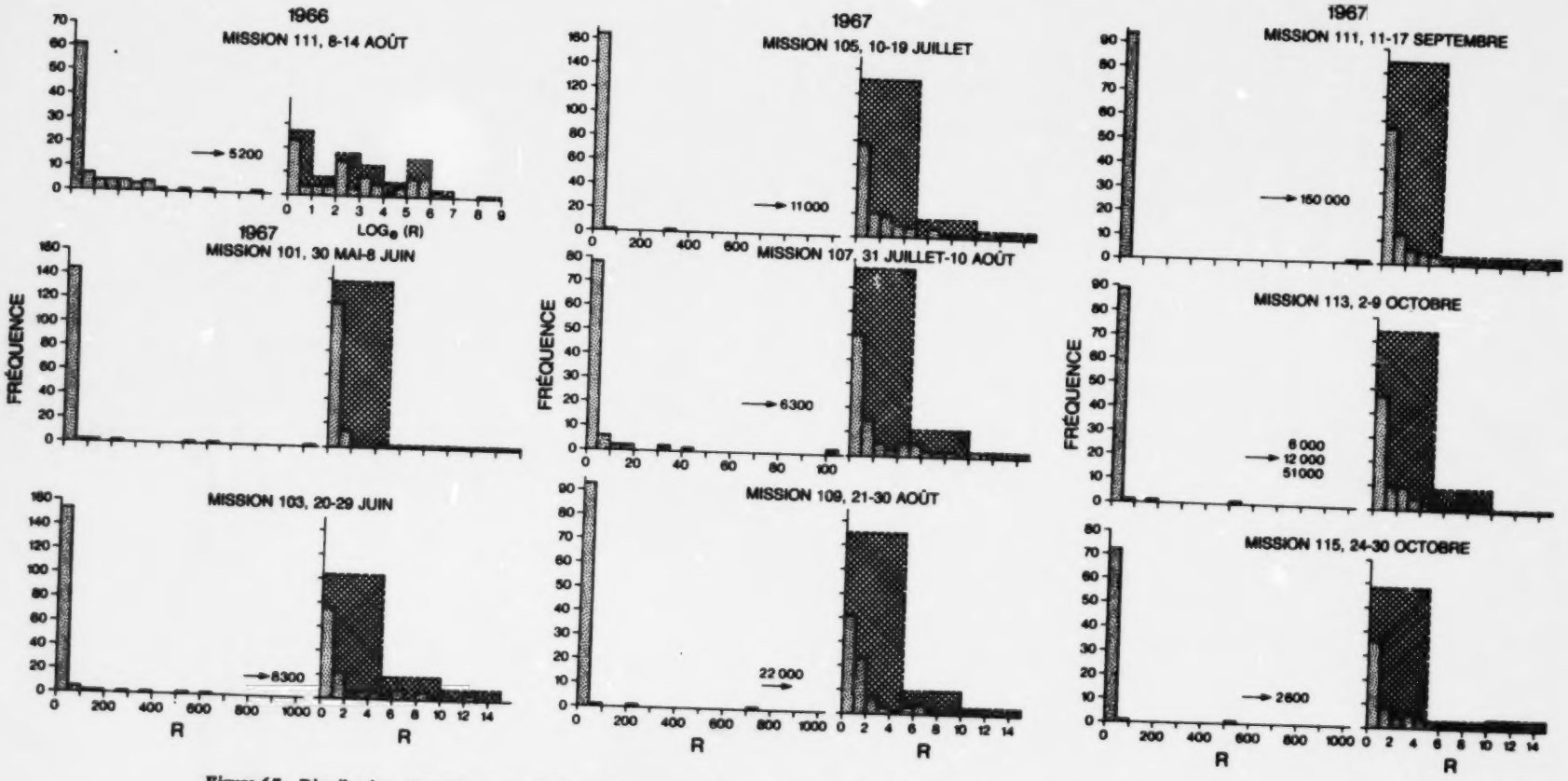



**Figure 67. Distributions des fréquences observées des concentrations de coliformes totaux (R) ou de $\log_e R$ présentées selon deux échelles.**

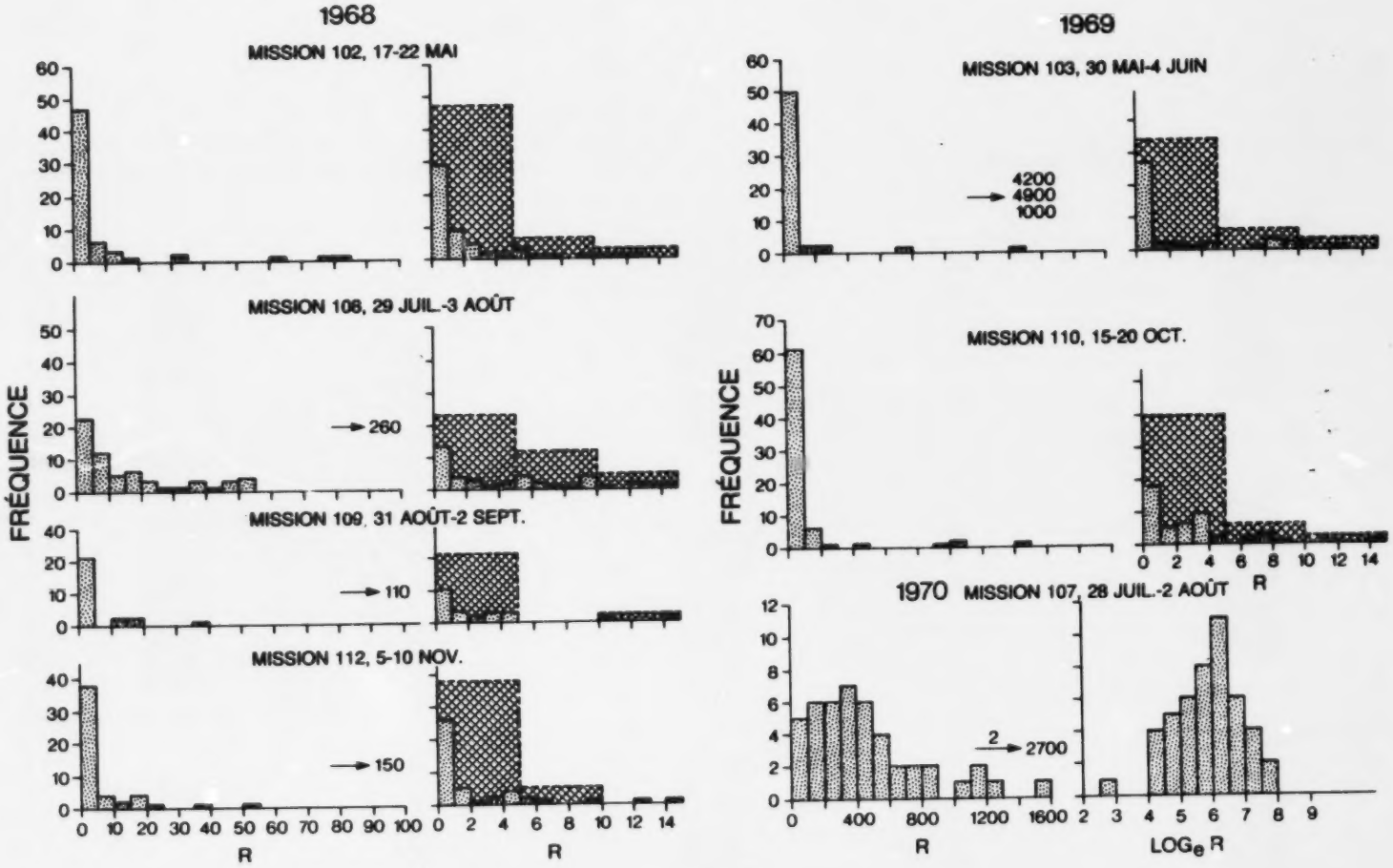



Figure 67 (suite)

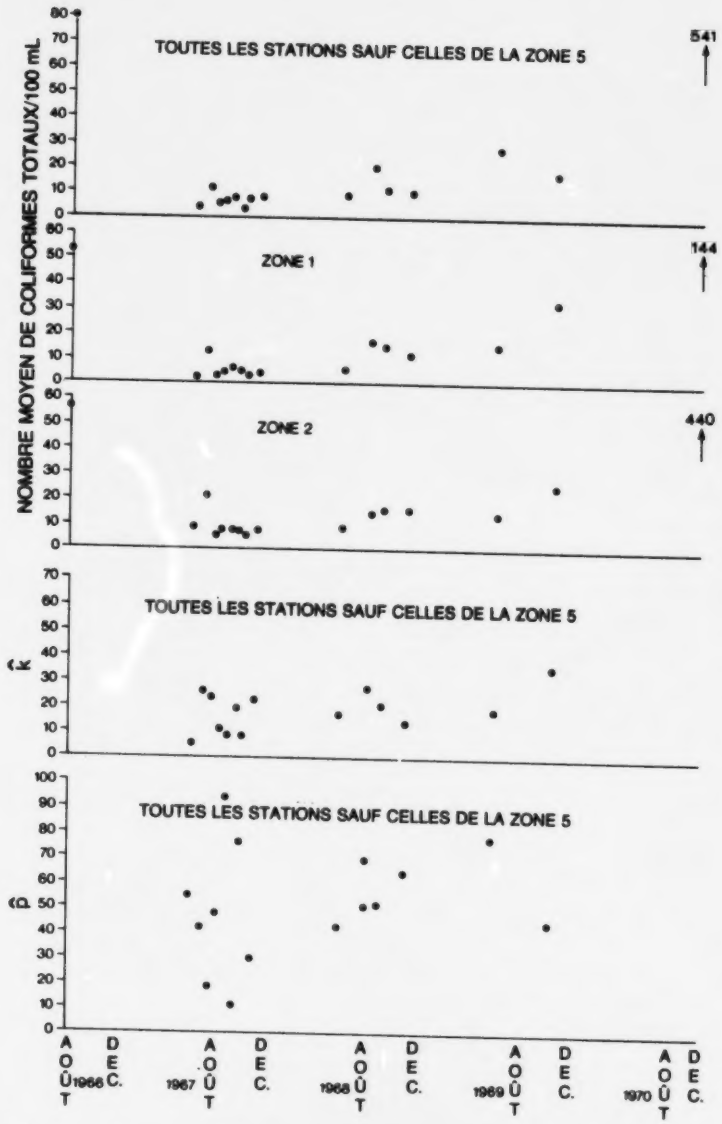



**Figure 68. Concentrations moyennes de coliformes totaux dex zones 1 à 4 et des zones 1 et 2, et estimations par maximum de vraisemblance des paramètres de la distribution binomiale négative pour chaque mission.**

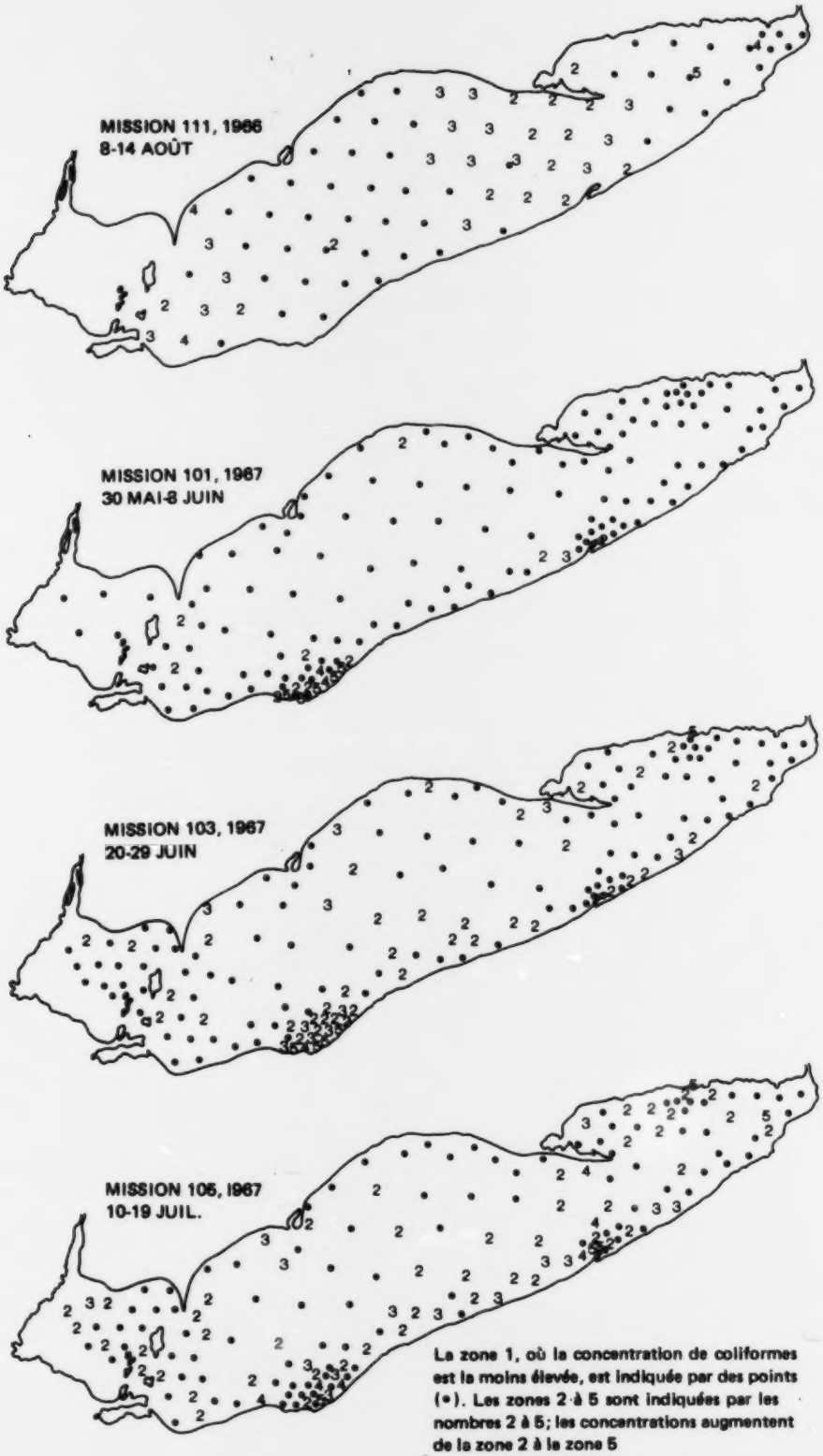
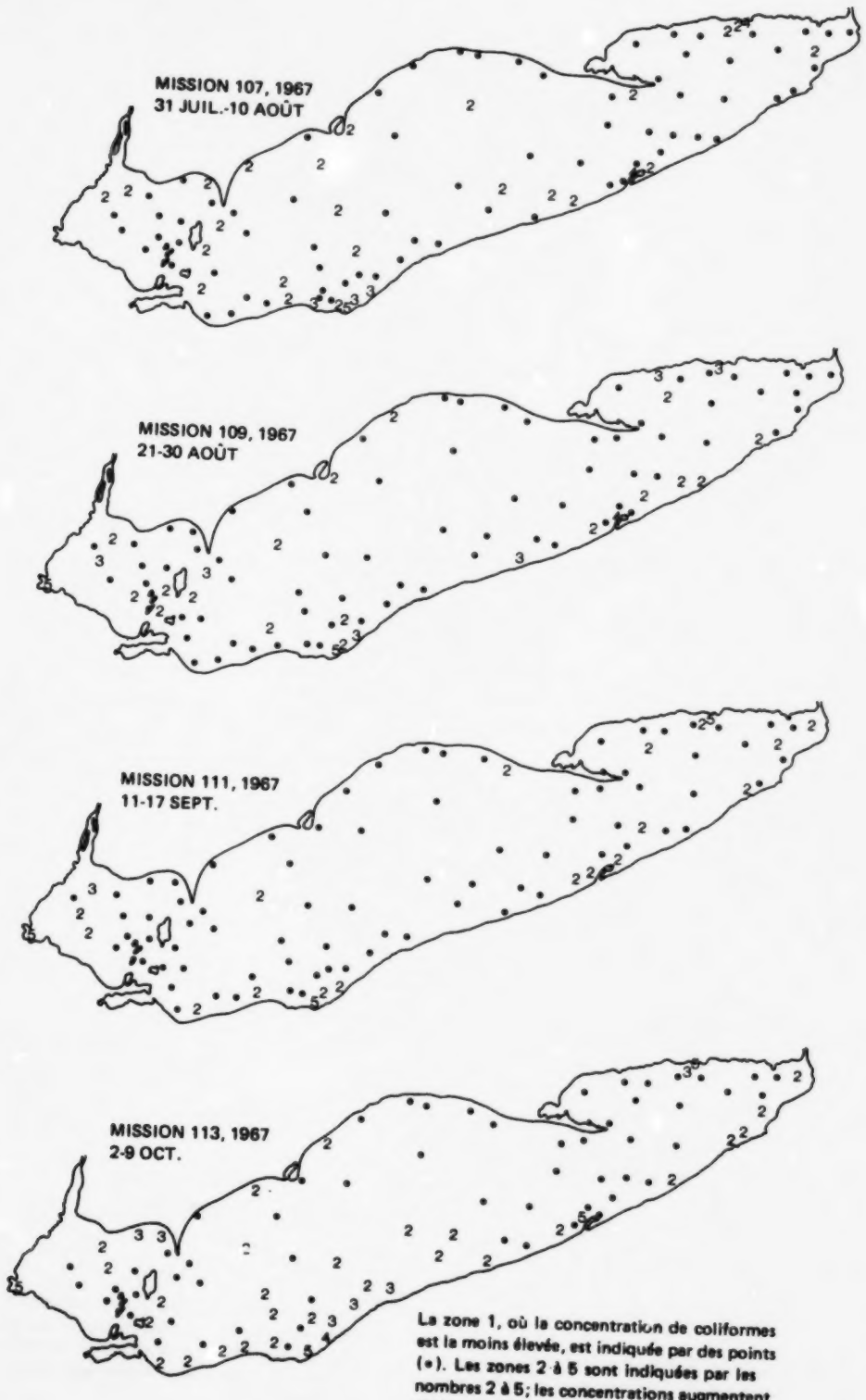
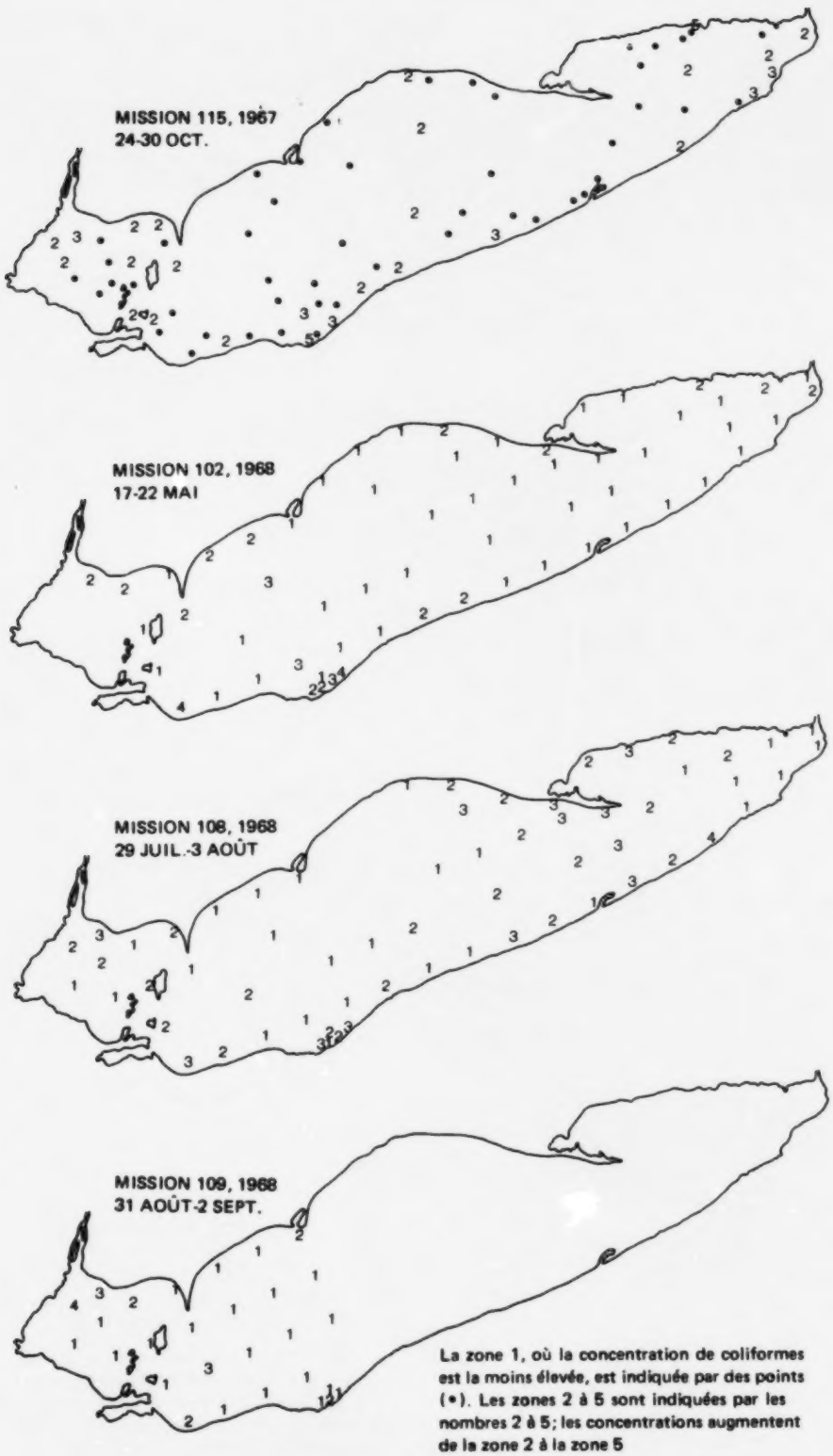



**Figure 69. Zones du lac Érié, selon les concentrations de coliformes totaux, par mission, de 1966 à 1970.**



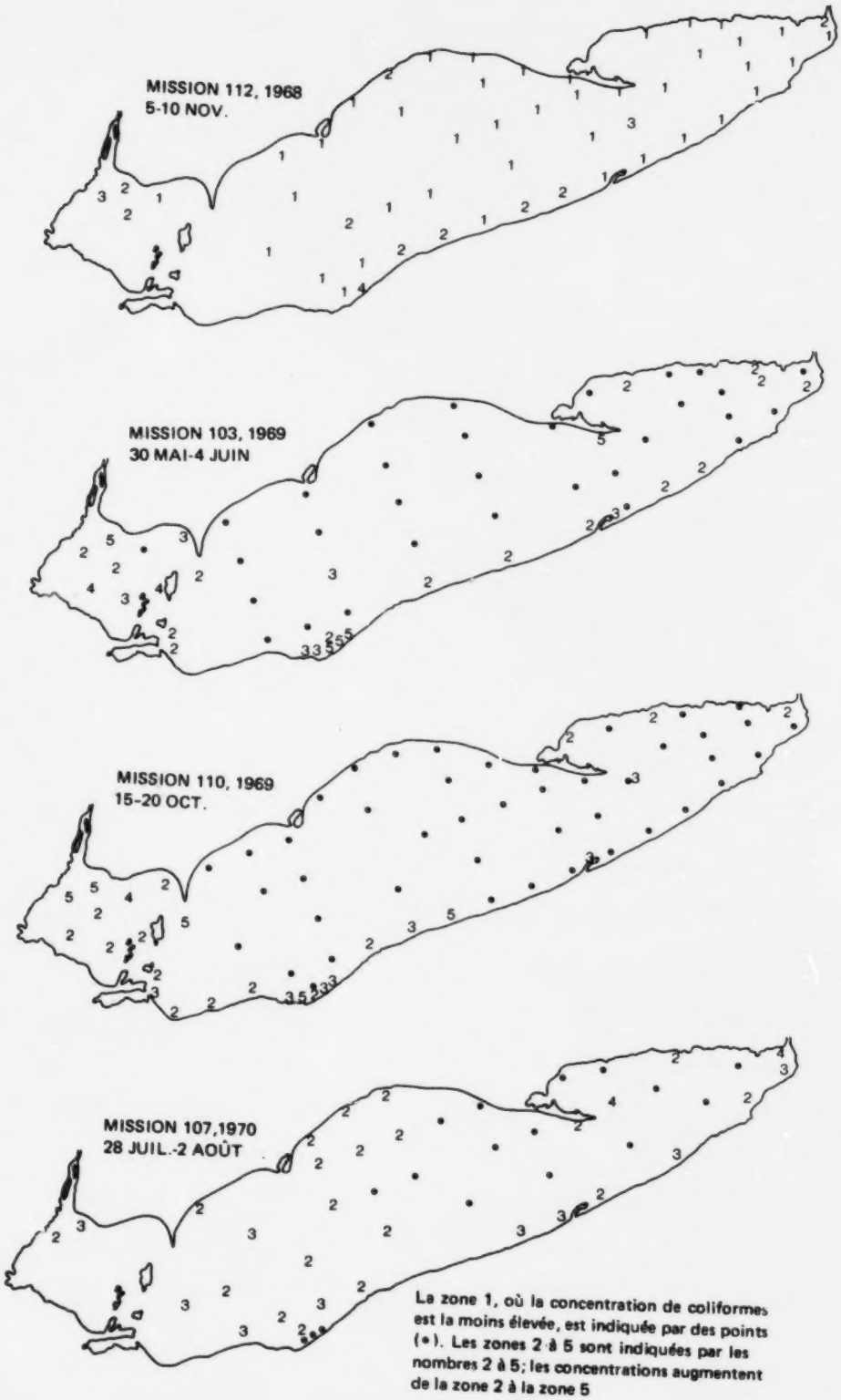

La zone 1, où la concentration de coliformes est la moins élevée, est indiquée par des points (•). Les zones 2 à 5 sont indiquées par les nombres 2 à 5; les concentrations augmentent de la zone 2 à la zone 5

**Figure 69 (suite)**




La zone 1, où la concentration de coliformes est la moins élevée, est indiquée par des points (•). Les zones 2 à 5 sont indiquées par les nombres 2 à 5; les concentrations augmentent de la zone 2 à la zone 5

**Figure 69 (suite)**




La zone 1, où la concentration de coliformes est la moins élevée, est indiquée par des points (•). Les zones 2 à 5 sont indiquées par les nombres 2 à 5; les concentrations augmentent de la zone 2 à la zone 5

**Figure 69 (suite)**

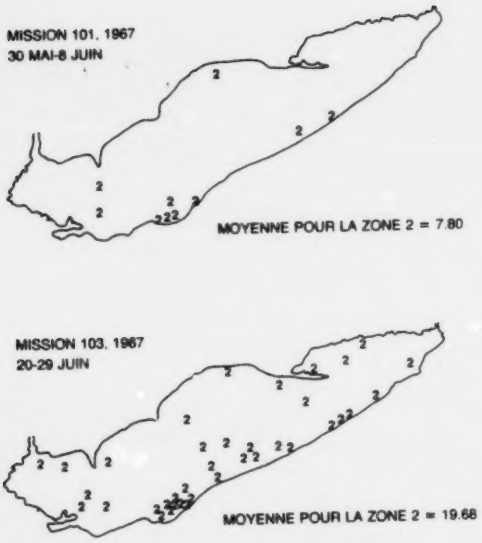



**Figure 70. Exemples de zones 2 caractéristiques pour des distributions de coliformes totaux de printemps et maximales.**

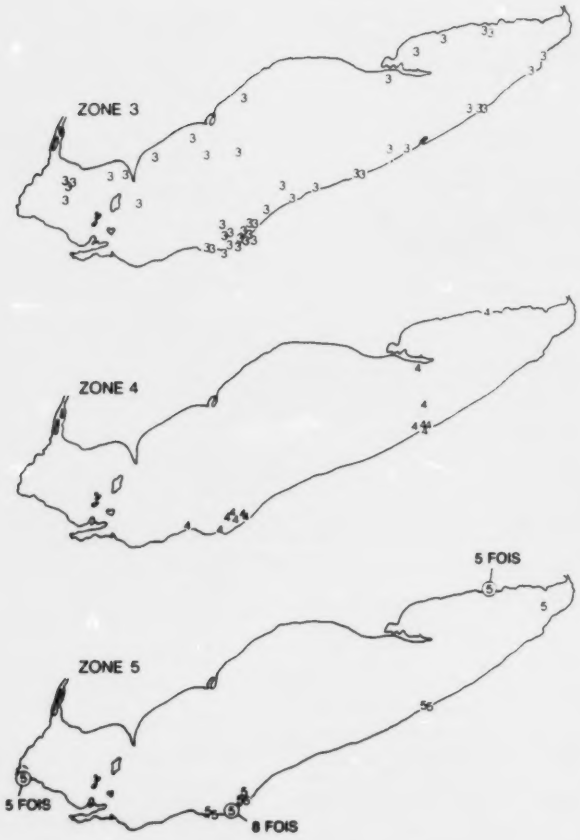



**Figure 71. Zones 3, 4 et 5 selon les concentrations de coliformes totaux regroupées pour toutes les missions de 1967.**

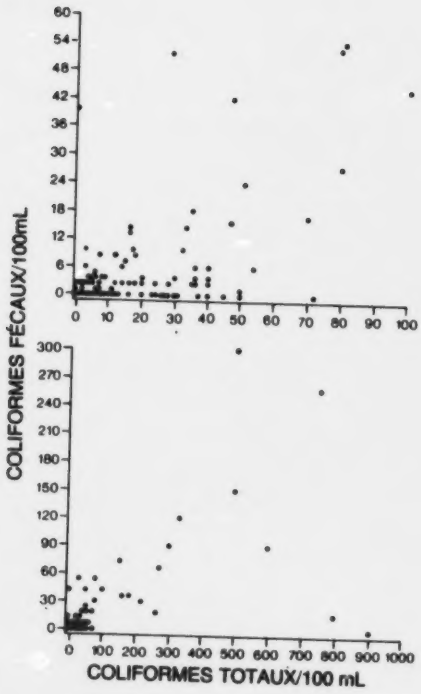



**Figure 72. Concentrations de coliformes fécaux en fonction de deux gammes de concentrations de coliformes totaux — missions de 1967 et de 1968.**

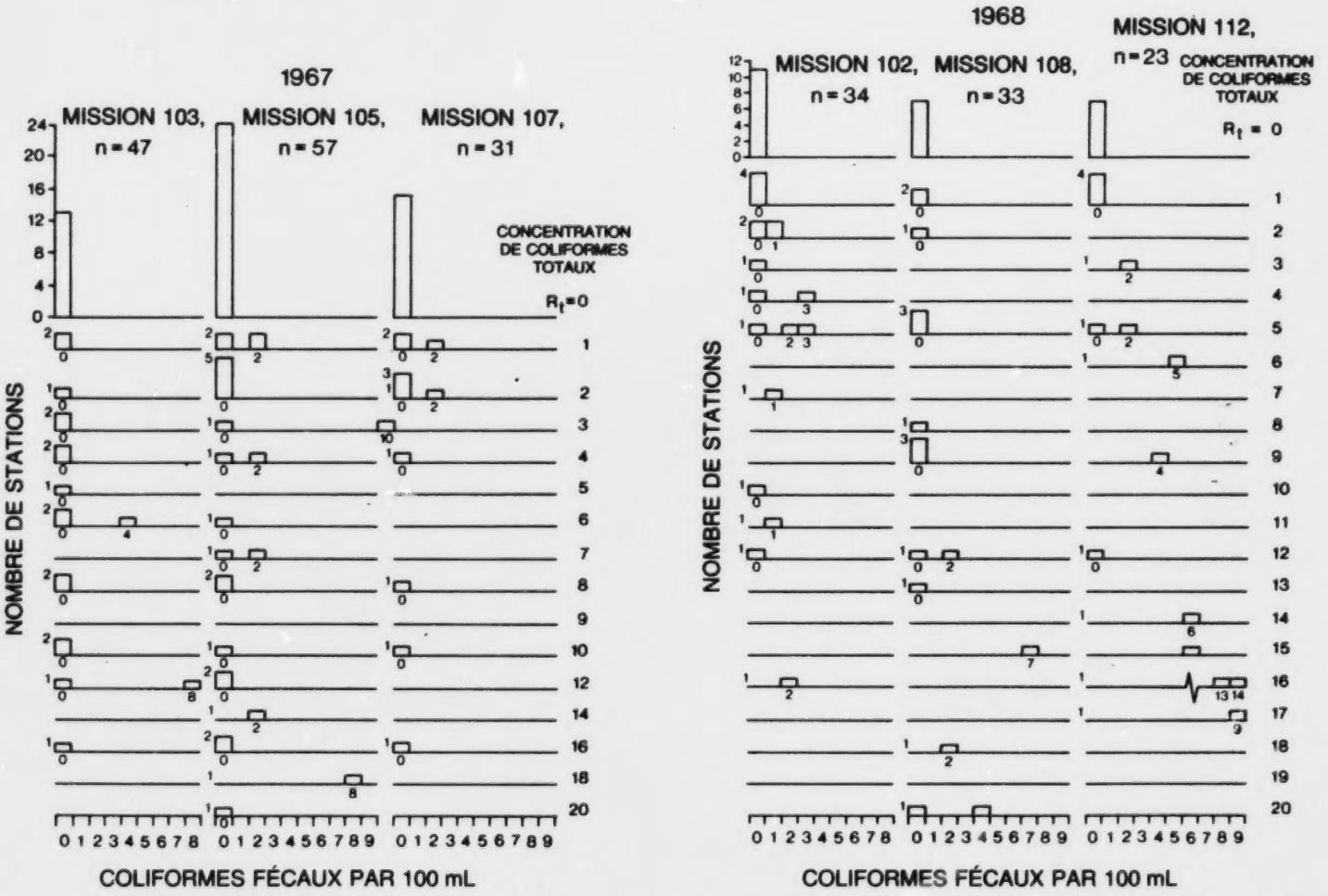



**Figure 73. Distributions des fréquences conditionnelles des concentrations de coliformes fécaux pour une concentration donnée de coliformes totaux ($R_t$) pour $R_t < 20$, lors de trois missions de 1967 et de trois missions de 1968 (n = nombre total de stations communes).**

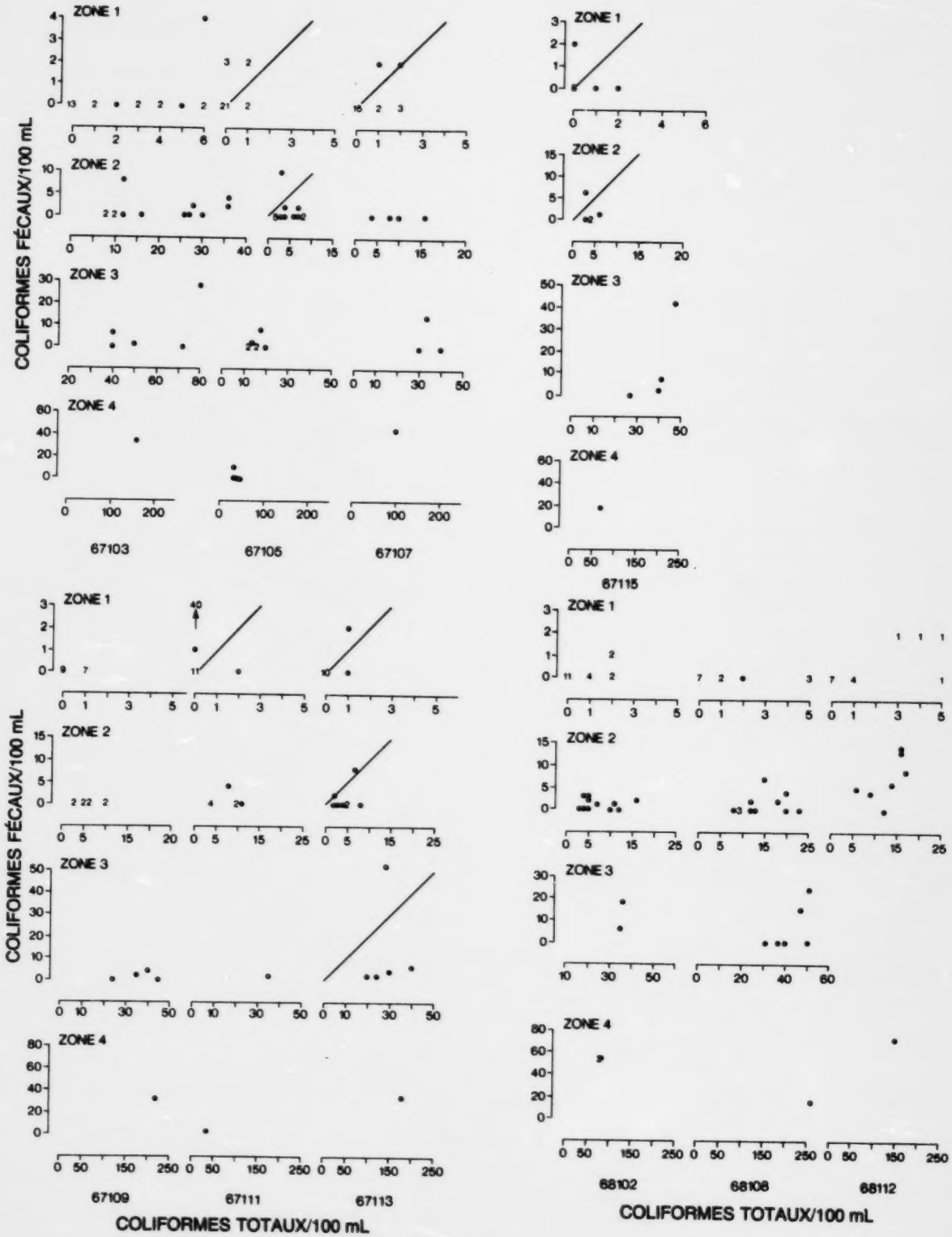



**Figure 74. Concentrations de coliformes fécaux en fonction des concentrations de coliformes totaux par zone et par mission pour 1967 et 1968.**

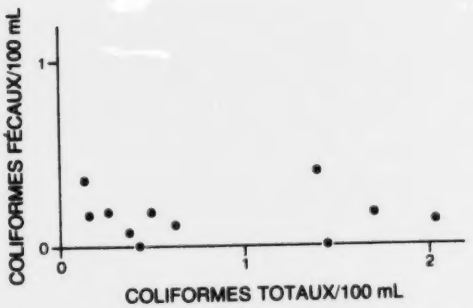



**Figure 75. Concentration moyenne de coliformes fécaux en fonction de la concentration moyenne de coliformes totaux dans la zone 1 pour les missions de 1967 à 1969.**

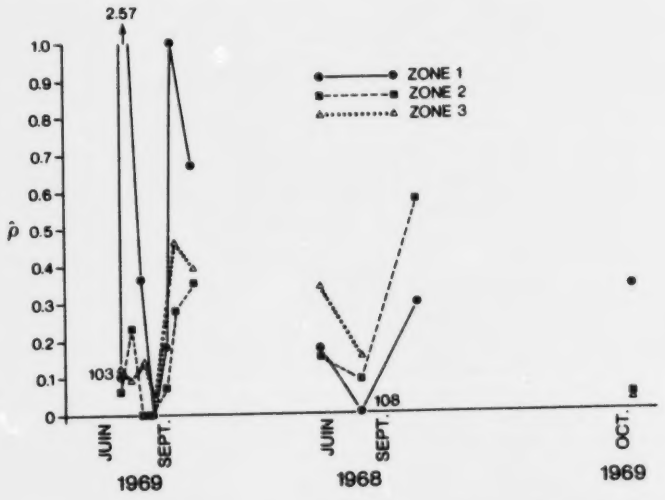



**Figure 76. Rapports de la concentration des coliformes fécaux à celle des coliformes totaux en fonction de la date des missions — zones 1, 2 et 3, données des stations communes.**

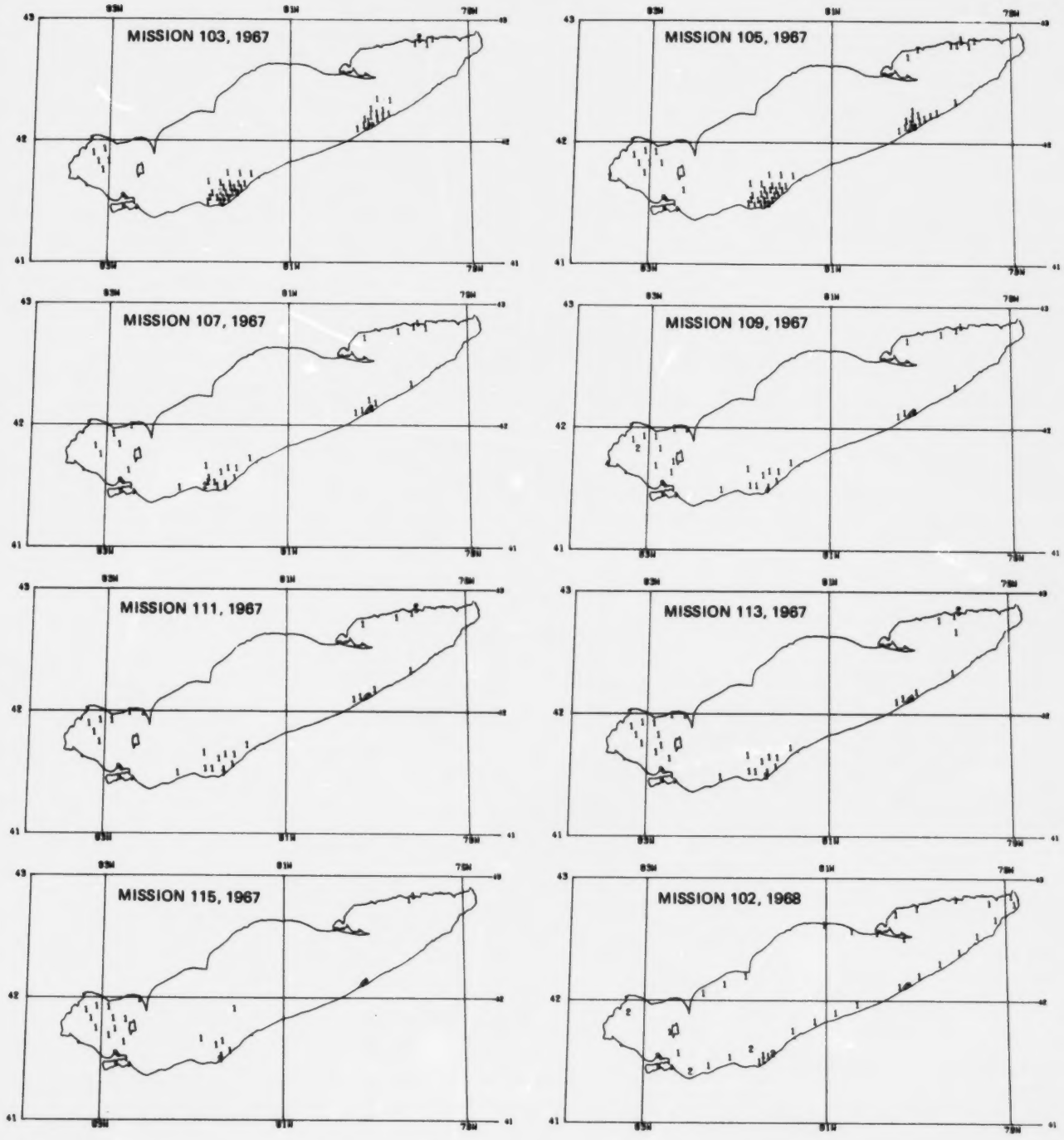
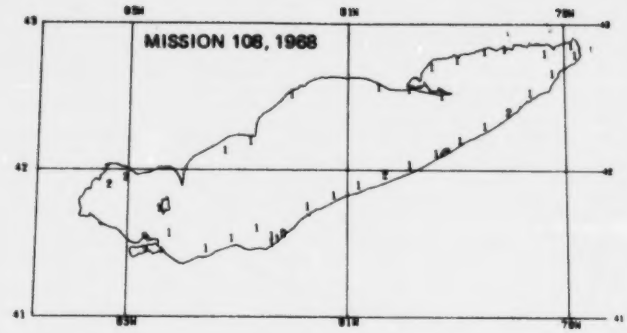
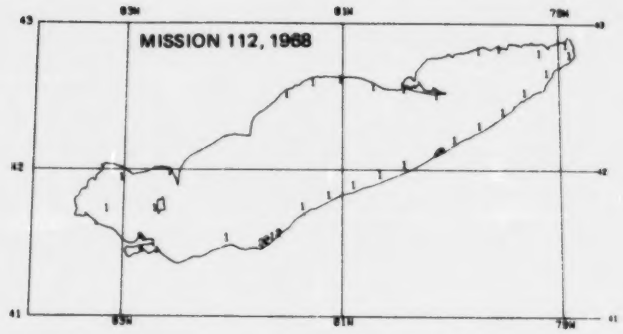
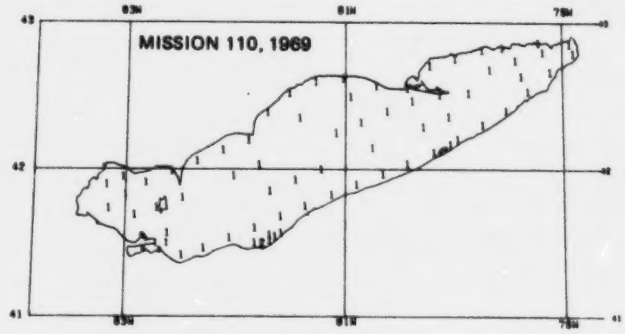
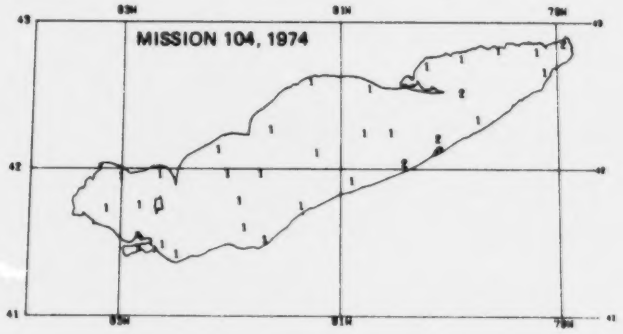


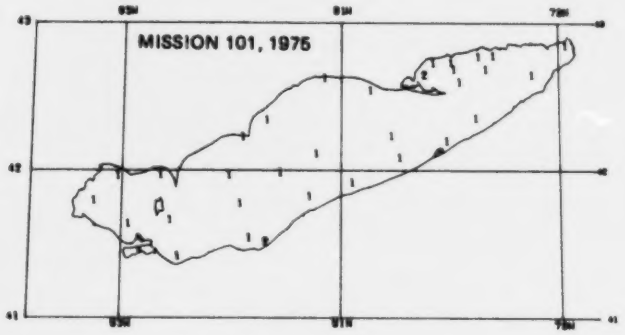

**Figure 77. Emplacements des stations où les concentrations de coliformes fécaux ont été déterminées.**

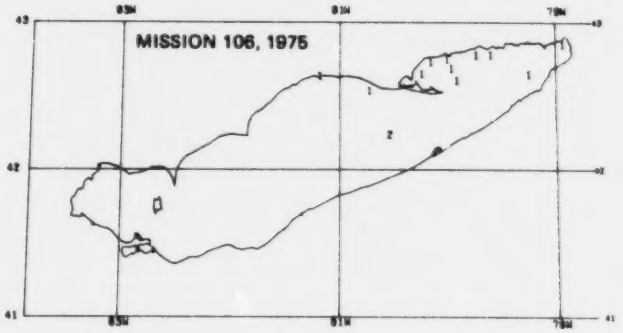


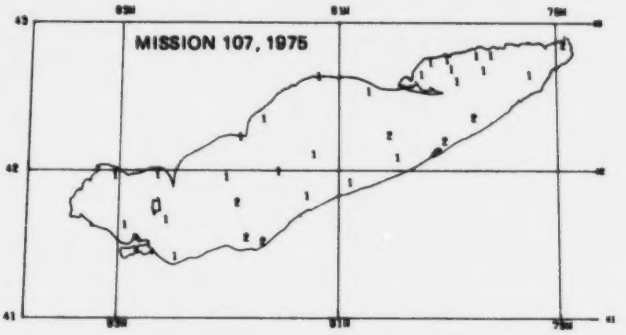




**Figure 77 (suite)**

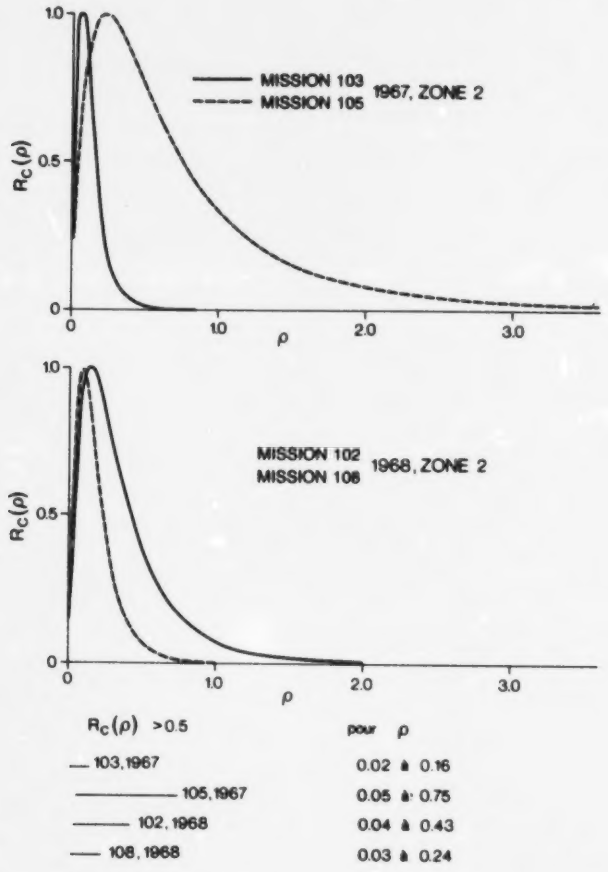



**Figure 78. Fonction de vraisemblance conditionnelle relative $R_c(\rho)$ pour le rapport des concentrations de coliformes fécaux et totaux lors de quatre missions.**

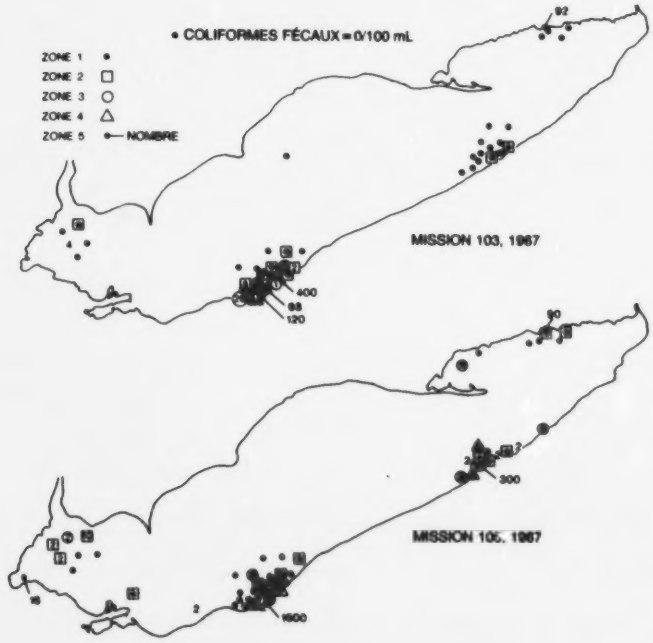



**Figure 79. Concentrations de coliformes fécaux dans les zones de coliformes totaux – deux exemples.**

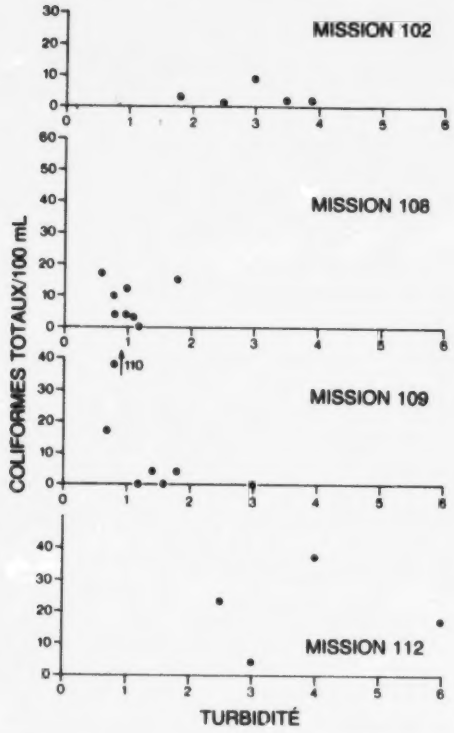



**Figure 80. Concentration de coliformes totaux en fonction de la turbidité dans la bassin ouest pendant les missions de 1968.**

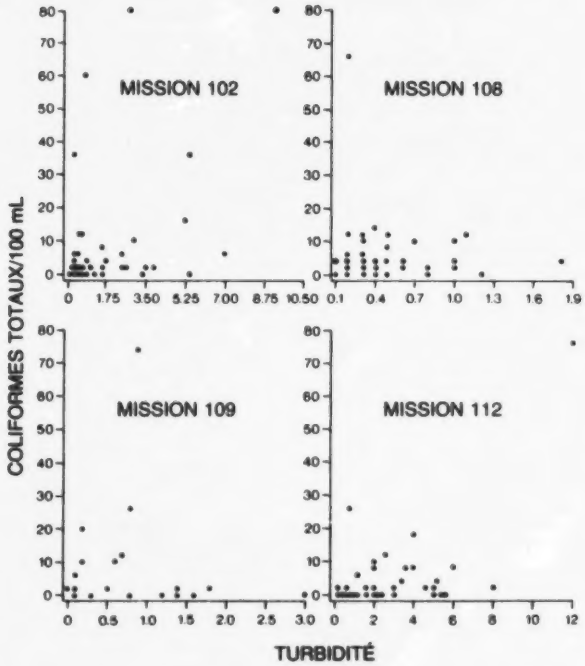



**Figure 81. Concentration de coliformes totaux en fonction de la turbidité lors des missions de 1968.**

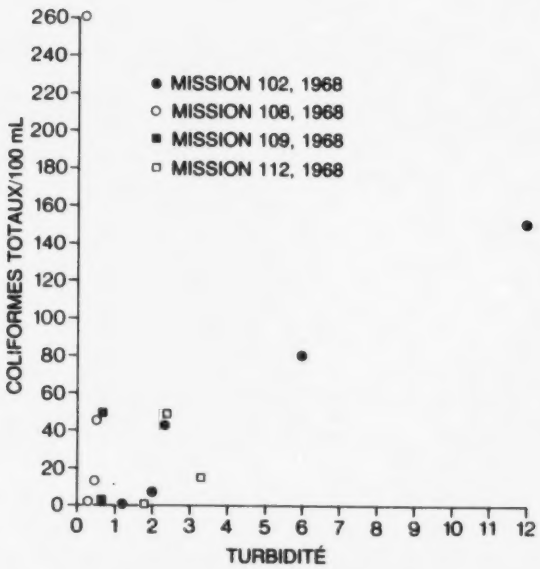



**Figure 82. Concentration moyenne de coliformes totaux en fonction de la turbidité moyenne lors des missions de 1968 aux stations communes des zones 1 à 4 déterminées à partir de l'ensemble des données sur les coliformes totaux.**

# Stratégie d'échantillonnage pour future cueillette de données

*par A.H. El-Shaarawi*

## INTRODUCTION

La qualité de l'eau d'un lac, ou de tout autre important plan d'eau, ne peut être évaluée que par échantillonnage. Plusieurs organismes s'intéressant à la qualité de l'eau, tels que le Centre canadien des eaux intérieures (CCEI), effectuent chaque année plusieurs relevés par échantillonnage. La première étape d'un relevé consiste à élaborer une stratégie d'échantillonnage où l'on précise: (i) les paramètres à mesurer, (ii) le nombre de stations de prélèvement, (iii) l'emplacement des stations et (iv) les profondeurs auxquelles les échantillons d'eau seront prélevés, ceci pour chacune des stations.

Plusieurs facteurs influent sur la précision des résultats obtenus. Le premier, et probablement le plus important est le choix du nombre de stations et de leur emplacement. Cela s'avère particulièrement vrai si la variable limnologique présente une importante variation spatiale au sein du lac. Un mauvais choix de l'emplacement des stations se traduira par des données ne représentant pas adéquatement l'état du lac et probablement par une représentation biaisée donc, des résultats inexacts. Il est donc primordial d'utiliser toute l'information disponible sur la variabilité spatiale lors de la planification des relevés.

Le deuxième facteur est le temps nécessaire au relevé. Ce facteur est important car les propriétés limnologiques de l'eau

dépendent de plusieurs variables environnementales, comme le vent, la pluie et le ruissellement, et la majorité de ces variables ne peuvent être modifiées ou prévues, surtout pour une longue période. Il est donc avantageux de minimiser les effets de ces variables en réduisant la durée des relevés. Il est évident que ce facteur a des effets opposés à ceux du facteur précédent, car plus le nombre de stations et plus la distance qui les sépare sont importants, plus il faut de temps pour effectuer les relevés.

Plusieurs autres facteurs influent sur la précision du relevé; il s'agit du soin et de la constance dans le prélèvement et l'analyse des échantillons d'eau, et dans l'enregistrement des résultats. Nous ne traiterons pas ici des effets de ces facteurs.

L'élaboration d'une stratégie d'échantillonnage exige de préciser les objectifs du relevé. Ces derniers, avec les contraintes financières et pratiques, peuvent être représentés par une expression mathématique qui spécifie le nombre et l'emplacement des stations de prélèvement.

On s'intéresse, dans le présent chapitre, à la façon d'utiliser l'information disponible sur la variabilité spatiale et temporelle des données limnologiques en vue d'élaborer une stratégie d'échantillonnage visant l'estimation de la valeur moyenne pondérée dans l'espace d'une seule variable limnologique. Les résultats présentés ici sont ceux d'une étude préliminaire du problème de la conception d'un relevé qui met l'accent sur quelques principes généraux. On suppose deux types de représentation spatiale. On suppose dans le premier cas, que le lac est formé d'une série de zones «homogènes» discrètes et, dans le second, que les valeurs de la variable limnologique peuvent être considérées comme celles d'une variable aléatoire continue dont la moyenne peut être exprimée sous forme d'une fonction continue (surface) de la position des stations d'échantillonnage. La généralisation à plusieurs variables s'avère

très difficile et doit s'inspirer à la fois de la statistique et de la limnologie; elle nécessite en outre une base de données supérieure à celle dont nous disposons pour le lac Erié. Nous ne traiterons donc pas de ce cas.

Les approches statistiques sont illustrées et traitées à l'aide de données sur le dénombrement des coliformes, la température et la teneur en chlorophylle a. On présente, dans la section suivante, un exemple hypothétique visant à illustrer les principes d'une stratégie d'échantillonnage. On suppose, de façon plus précise, que l'on cherche à estimer la concentration bactérienne moyenne d'un lac, mais le traitement peut s'appliquer à toute autre variable.

## ILLUSTRATION DES PRINCIPES DE L'ECHANTILLONNAGE

Supposons que le relevé a pour objet d'obtenir une estimation du nombre de bactéries, $\mu$, par unité de volume. La méthode normale consiste à prélever un certain nombre d'échantillons d'eau et à estimer $\mu$ par la moyenne arithmétique ou géométrique des données. Il est important de noter que la valeur d'un tel calcul de la moyenne est fonction de l'endroit où les échantillons sont prélevés. On peut, pour illustrer ce point, prendre comme exemple un lac idéal dont la forme est celle d'une boite de longueur $\ell$, de largeur W et de profondeur D, où la distribution des bactéries est uniforme à la verticale. Supposons de plus que le nombre de bactéries par unité de volume, la densité, obéit à la relation linéaire suivante:

$$\mu_\ell = \alpha + \beta_\ell \quad 0 \leq \ell \leq L \tag{7.1}$$

où $\mu_\ell$ est la densité bactérienne en un point situé à $\ell$ unités de l'extrémité du lac qui sert de point d'origine pour la mesure de toutes les distances, et où $\alpha$ et $\beta$ sont des constantes

inconnues. L'équation 7.1 intégrée sur tout le lac et divisée par le volume du lac devient:

$$\mu = \alpha + \beta L/2 \qquad (7.2)$$

Supposons que n échantillons ont été prélevés dans le lac à des distances $\ell_1$, $\ell_2 \ldots \ell_n$ et qu'il n'y ait aucune erreur aléatoire possible. Dans ces conditions, cette stratégie d'échantillonnage nous donne $\mu_{\ell 1}, \mu_{\ell 2}, \ldots \mu_{\ell n}$ qui représentent les dénombrements bactériens observés. Quand la moyenne $\bar{\mu} = (\mu_{\ell 1} + \mu_{\ell 2} + \ldots + \mu_{\ell n})/n$ sert à estimer $\mu$, on obtient, par l'équation 7.1, le biais:

$$B = \mu - \bar{\mu} = \beta(\frac{L}{2} - \bar{\ell})$$

La valeur de B représente le biais découlant de l'estimation de $\mu$ par ce plan d'échantillonnage particulier. Le biais est nul quand $\beta = 0$; ce qui suppose que le lac est totalement homogène ou que la stratégie d'échantillonnage est telle que $\bar{\ell} = L/2$. On obtient un biais positif maximum de $\beta L/2$ lorsque $\ell_1 = \ell_2 = \ldots = \ell_n = 0$ et un biais négatif maximum de $-\beta L/2$ lorsque $\ell_1 = \ell_2 = \ldots = \ell_n = L$. On voit que l'emplacement des stations d'échantillonnage joue un rôle très important dans l'estimation sans biais de $\mu$ quand le lac est très hétérogène.

La condition $\bar{\ell} = L/2$ ne garantit pas un bon plan d'échantillonnage, car il est possible de prélever tous les échantillons d'eau au point $L/2$ de façon à obtenir une estimation non biaisée de $\mu$. Il faut donc introduire un autre critère: on suppose que les valeurs de $\mu_\ell$ ne sont pas observées exactement, mais qu'elles sont sujettes à des erreurs aléatoires de moyenne 0 et de variance $\sigma^2$. Supposons, pour être plus précis, qu'il est impossible d'observer $\mu_\ell$ mais qu'il est possible d'observer

$C_\ell = \mu_\ell + \varepsilon_\ell$ où $\varepsilon_\ell$ est une variable aléatoire de moyenne 0 et de variance $\sigma^2$. On peut, à partir de ces hypothèses, réécrire l'équation 7.1 de la façon suivante:

$$C_\ell = \alpha + \beta\ell + \varepsilon_\ell \tag{7.3}$$

Connaissant $C_{\ell 1}$, $C_{\ell 2} \ldots C_{\ell n}$, les valeurs de $\alpha$ et $\beta$ peuvent être estimées, par régression par exemple. Soit $\hat{\alpha}$ et $\hat{\beta}$ les estimations de $\alpha$ et $\beta$. La valeur estimée de $\mu$ est alors $\hat{\mu} = \hat{\alpha} + \hat{\beta}L/_2$. Il est facile de montrer que la variance de $\hat{\mu}$ ($\alpha$ et $\beta$ sont estimées par régression) est:

$$\text{var}\ (\hat{\mu}) = \{\frac{\sigma^2}{n} 1 + \frac{(\frac{L}{2} - \bar{\ell})^2}{s^2_\ell}\} \tag{7.4}$$

où $s^2_\ell = \frac{1}{n} \sum_{i=1}^{n} (\ell_i - \bar{\ell})^2$. L'équation 7.4 donne la précision de l'estimation de $\mu$. Plus la valeur de var $(\hat{\mu})$ est importante, plus la précision diminue. La précision maximale est donc obtenue quand $\frac{L}{2} - \bar{\ell} = 0$ et $s^2_\ell \neq 0$. De fait, il est nécessaire, pour maximiser la précision, de choisir le plan qui maximise $S^2_\ell$ et minimise $\frac{L}{2} - \bar{\ell}$. On y parvient en n'utilisant que deux valeurs de $\ell_i$, à savoir 0 et L, et en rattachant chaque moitié de l'échantillon à chaque extrémité du lac.

Cet exemple montre qu'il est possible de choisir une stratégie d'échantillonnage appropriée à l'estimation de la densité bactérienne d'un lac lorsque l'on dispose de certaines informations sur la variabilité spatiale des dénombrements bactériens, sujet qui sera traité dans la prochaine section.

## STRATEGIE D'ECHANTILLONNAGE D'UN LAC DIVISE EN ZONES

Supposons que l'on divise un lac en k zones homogènes en appliquant, par exemple, la méthode de El-Shaarawi et coll. (1981) aux données chronologiques. Supposons aussi que le but du relevé est d'estimer la densité bactérienne du lac pondérée en fonction de la surface. Selon l'hypothèse d'homogénéité, le nombre de bactéries $r_{ji}$ du $j^{ième}$ échantillon ($j = 1,2...n_i$) provenant de la $i^{ième}$ zone ($i = 1,2...k$) obéit à la loi de Poisson,

$$P(r_{ji}) = e^{-\mu_i} \mu_i^{r_{ji}/r_{ji}}$$

de moyenne $\mu_i$. Le paramètre $\mu_i$ représente la densité bactérienne réelle mais inconnue dans la $i^{ième}$ zone; son estimation est $\bar{r}_i$, la moyenne arithmétique de $r_{1i}, r_{2i} \cdots r_{ni}$. Soit n le nombre de stations à échantillonner pendant le relevé. La densité moyenne surfacique est donc

$$\bar{r}_\omega = \sum_{i=1}^{k} A_i \bar{r}_i / \sum_{i=1}^{k} A_i$$

où $A_i$ est la surface de la $i^{ième}$ zone. La variance de $\bar{r}_\omega$ est estimée par

$$S^2_{\bar{r}_\omega} = \left( \sum_{i=1}^{k} A_i^2 \bar{r}_i / n_i \right) / \left( \sum_{i=1}^{k} A_i \right)^2 \qquad (7.5)$$

La répartition optimale des n stations dans k zones est donnée par les valeurs $n_1, n_2 \cdots n_k$ qui minimisent var $(\bar{r}_\omega)$ tout en respectant la condition voulant que $n_1 + n_2 + \ldots n_k = n$. On peut facilement montrer que la valeur de $n_i$ est donnée par

$$n_i = n(A_i\bar{r}_i)/\sum_{i=1}^{K} A_i\bar{r}_i \quad (i = 1,2...k) \qquad (7.6)$$

Cette formule montre que le nombre de stations affectées à la $i^{ième}$ zone est proportionnel à la surface de cette zone multipliée par la densité bactérienne moyenne dans cette même zone. Il faut noter que lorsque les $r_i$ sont presque égaux, la répartition dans les zones est alors approximativement proportionnelle à la surface de ces dernières. En substituant la valeur de $n_i$ de l'équation 7.6 dans l'équation 7.5, on obtient une estimation de la valeur minimale de la variance $\bar{r}_w$:

$$S^2_2 = (\sum_{i=1}^{K} A_i\bar{r}_i)/n \sum_{i=1}^{K} A_i \qquad (7.7)$$

On suppose généralement, pour déterminer l'intervalle de confiance de la densité bactérienne selon la répartition optimale, que $\bar{r}_\omega$ est normal; cela est justifié par le théorème de la limite centrale avec la variance donnée par 7.7. Le nombre de stations n où il faut effectuer des prélèvements pour obtenir une différence préétablie $\bar{d} = (r_\omega - \mu)$ entre la densité estimée $r_\omega$ et la densité réelle mais inconnue, $\mu$, peut être déterminée. Ainsi, la valeur de n nécessaire pour déceler la différence d au niveau de signification de 5 % est:

$$n \leq (1.96)^2(\sum_{i=1}^{k} A_i\bar{r}_i/\sum_{i=1}^{k} A_i)/d^2 \qquad (7.8)$$

## APPLICATIONS

Les données traitées par El-Shaarawi et coll. (1981) sont utilisées pour illustrer les méthodes présentées par El-Shaarawi et Esterby (1981). Il s'agit de dénombrements de coliformes obtenus par

la technique de filtration sur membrane avec incubation de 24 h à 35°C sur milieu M. Endo (Difco). Des analyses bactériologiques ont porté sur des échantillons d'eau prélevés en 1968 en diverses stations du lac Erié lors des quatre missions suivantes: (i) 17 au 24 mai, (ii) 29 juillet au 3 août, (iii) 31 août au 3 septembre, et (iv) du 4 au 10 novembre. El-Shaarawi et coll. (1981) ont montré que les dénombrements bactériens du lac présentaient une très forte hétérogénéité spatiale; le lac a donc été divisé en plusieurs zones homogènes (voir figure 4 dans El-Shaarawi et coll., 1981). Les résultats de l'étude précédente El-Shaarawi sont utilisés dans la présente section pour: (i) répartir de façon optimale les stations dans les zones, en supposant un nombre de stations égal à celui du relevé de 1968 et (ii) pour déterminer le nombre de stations à échantillonner ainsi que leur distribution dans le lac, ceci pour obtenir exactement la même efficacité que celle du plan d'échantillonnage de 1968. La différence entre ce nombre et le nombre correspondant de stations utilisées en 1968 renseigne sur l'importance des ressources économisées par l'adoption de la méthode de répartition optimale qui, bien sûr, aurait pu être appliquée si l'on avait disposé d'informations relatives à la variabilité bactérienne du lac avant d'effectuer les relevés de 1968.

On pourra trouver la moyenne arithmétique $\bar{r}_a$ et la moyenne pondérée en fonction de la surface $\bar{r}_\omega$ de la densité bactérienne aux colonnes 2 et 3, respectivement, du tableau 66. La variance réelle $r_\omega$ selon ce plan est donnée par l'équation 7.5; les valeurs sont présentées dans la colonne 4. On trouvera à la colonne 5 l'estimation correspondante de la variance après répartition optimale selon l'équation 7.7. Le rapport de la variance optimale à la variance réelle, présenté dans la colonne 6, indique l'efficacité du plan de 1968. Ainsi, l'efficacité minimale est de 0.387 et correspond à la mission effectuée du 17 au 24 mai, tandis que l'efficacité maximale est de 0.720 et correspond à la mission du 20 juillet au 3 août. Une autre façon de mesurer l'efficacité

consiste à déterminer le nombre de stations à échantillonner lorsqu'on fait appel à la répartition optimale pour obtenir la même variance que celle obtenue en 1968. On suppose que la variance $s^2_2$ de $\bar{r}_\omega$ de l'équation 7.7 est égale à la variance réelle observée et il reste à trouver la valeur de n pour chacune des zones. Ces valeurs sont présentées à la colonne 8 du tableau 66 par $\tilde{n}$ qui peuvent être comparées au nombre réel de stations n utilisées en 1968 (colonne 7).

Une fois le nombre de stations précisé, la formule 7.6 sert à déterminer le nombre de stations qui devraient être affectées à chacune des zones. Supposons tout d'abord que le nombre des stations à échantillonner est le même que celui utilisé en 1968; cela permet de comparer la répartition optimale et la répartition réelle. Deuxièmement, supposons que le nombre de stations utilisées permet d'obtenir la même efficacité que celle de 1968. Les résultats de ces calculs, basés sur les zones définies dans l'étude de El-Shaarawi et coll. (1981), sont présentés au tableau 67. Les colonnes n donnent la répartition de 1968, les colonnes $\hat{n}$ la répartition optimale et les colonnes $\tilde{n}$ la répartition d'efficacité égale à celle de 1968. On note en général, dans le plan optimal (tableau 67), une réduction du nombre de stations dans la première zone et une augmentation de ce nombre dans les zones à densité bactérienne élevée, lorsqu'on compare n et $\hat{n}$.

On indique au tableau 68 et à la figure 83 le nombre de stations à échantillonner pour déceler une différence relative préétablie de $d/\bar{r}_\omega$ avec un niveau de signification de 5 %. On note que le nombre de stations décroit très rapidement en fonction de $d/\bar{r}_\omega$ et que la mission du 17 au 24 mai ressemble à celle du 20 juillet au 3 août, les résultats des deux autres missions étant semblables. On peut aussi voir à la figure ou au tableau que 25 stations suffisent pour déceler une différence relative de 0.15 dans chacune des quatre périodes étudiées, mais ce nombre serait inférieur à certaines dates.

L'application des méthodes aux données de 1968 du lac Erié montre que le plan utilisé n'était pas efficace. Cela s'explique par l'importante variabilité spatiale du lac qui fait qu'il faudrait plus de stations dans les zones très variables. En outre, il est possible de réduire de beaucoup le nombre des stations dans le prochain plan d'échantillonnage et d'obtenir la même efficacité qu'en 1968. On s'est aussi intéressé à l'évaluation du nombre de stations nécessaires pour déceler une différence relative donnée; 25 stations sont suffisantes pour déceler une différence relative de 0.15.

## GENERALISATION

On traite ici de la généralisation du plan d'échantillonnage quand le lac est divisé en k zones homogènes. Cette généralisation doit tenir compte de deux facteurs: le premier est le fait que la distribution de la variable en question est inconnue et le second est le coût de cueillette des données. Supposons que les données chronologiques permettent d'obtenir une estimation de l'écart-type dans chaque zone $s_1, s_2 \ldots s_k$. Supposons de plus que nous voulions estimer la moyenne, pondérée en fonction de la surface, de la variable limnologique. Alors, tout plan d'échantillonnage qui affecte $n_1$ stations à la $1^{re}$ zone, $n_2$ stations à la $2^{ième}$ zone ... $n_k$ stations à la $k^{ième}$ zone, mène à l'estimation de la moyenne, pondérée en fonction de la surface, $\bar{x}_\omega$ suivante:

$$\bar{x}_\omega = \left(\sum_{i=1}^{\kappa} A_i \bar{x}_i\right) / \sum_{i=1}^{\kappa} A_i \qquad (7.9)$$

où $\bar{x}_i$ est la moyenne de $n_i$ observations dans la $i^{ième}$ zone ($i = 1,2 \ldots \kappa$). Supposons aussi que la visite d'une station d'échantillonnage de la $i^{ième}$ zone coûte $c_i (i = 1,2 \ldots \kappa)$. On présume donc que le coût d'une observation est le même pour toutes les stations d'échantillonnage d'une même zone, mais qu'il varie

d'une zone à l'autre. Le coût total de la cueillette des données est donc:

$$C = (n_1c_1 + n_2c_2 + \ldots + n_\kappa c_\kappa) \tag{7.10}$$

La variance estimée de $\bar{x}_\omega$ est

$$s^2 = \text{var}\ (\bar{x}_\omega) = \sum_{i=1}^{\kappa} A_i^2\ (s_i^2/n_i)/(\sum_{i=1}^{\kappa} A_i)^2 \tag{7.11}$$

Supposons que le coût total et le nombre de stations sont constants:

$$n = n_1 + n_2 + \ldots + n_\kappa \tag{7.12}$$

Le problème à résoudre consiste donc à estimer $n_i$ $(i = 1,2\ldots\kappa)$ de façon à minimiser le terme de l'équation 7.11 tout en respectant les conditions 7.10 et 7.12. Il existe une solution numérique, mais en négligeant la condition 7.12 on obtient l'expression de $n_i$ suivante:

$$n_i = \frac{CA_i.s_i}{(\sum A_i s_i)\sqrt{c_i}},\ (i = 1,2\ldots\kappa) \tag{7.13}$$

où l'on voit que le nombre de stations affectées à la $i^{\text{ième}}$ zone croit à mesure que l'écart-type et la surface de la zone augmentent et que le coût diminue. En substituant cette expression de $n_i$ dans 7.11, on obtient la valeur minimale de la variance selon la condition 7.10. La variance optimale peut alors servir à estimer le nombre de stations à échantillonner pour obtenir une précision pré-déterminée suivant la démarche de la section précédente.

## STRATEGIE D'ECHANTILLONNAGE POUR UNE CONCENTRATION A FONCTION CONTINUE

On suppose ici que les valeurs de la variable limnologique peuvent être représentées par une fonction continue à deux coordonnées $x_1$ et $x_2$ qui définissent la position de la station d'échantillonnage dans le lac ($x_1$ est la longitude et $x_2$ est la latitude):

$$y = f(x_1, x_2) + \varepsilon \tag{7.14}$$

où f représente la forme présumée de la fonction (surface) et $\varepsilon$ est une variable aléatoire de moyenne 0 et de variance $\sigma^2$. El-Shaarawi et Esterby (1981) ont utilisé, pour représenter certaines variables limnologiques, une surface quadratique de la forme:

$$f(x_1,x_2) = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_{12} x_1 x_2 + \alpha_{11} x_1^2 + \alpha_{22} x_2^2 = \underline{\alpha}\,\underline{a} \tag{7.15}$$

où $\underline{\alpha}$ est le vecteur des paramètres inconnus $\alpha_1$, $\alpha_2$, $\alpha_{12}$, $\alpha_{11}$, $\alpha_{22}$ et $\underline{a}$ est le vecteur 1, $x_1$, $x_2$, $x_1x_2$, $x^2_1$ et $x^2_2$. Soit $\hat{\underline{\alpha}}$ la valeur estimée de $\underline{\alpha}$ et $\hat{s}^2$ la valeur de $\sigma^2$ estimée à partir de données chronologiques. Dans ces conditions, l'objectif est d'utiliser le modèle 7.15 et l'estimation des paramètres pour élaborer une stratégie appropriée d'estimation de la surface. La moyenne pondérée de la variable limnologique est donnée par

$$\bar{y}_\omega = \iint_Q f(x_1,x_2)dx_1dx_2 \Big/ \iint_Q dx_1dx_2 \tag{7.16}$$

où Q représente la surface du lac. En substituant l'expression 7.15 dans 7.16 et intégrant le résultat, on obtient l'estimation $\bar{y}_\omega$ en fonction de $\underline{\alpha}$.

Quel que soit le plan d'échantillonnage, le prélèvement se fera en plusieurs endroits et l'on obtiendra l'estimation $\underline{\alpha}^*$ de $\underline{\alpha}$. La variance de $\alpha^*$ est une fonction des valeurs de $x_1$ et $x_2$; la variance de $\bar{y}_\omega$ sera donc aussi une fonction de $x_1$ et $x_2$. Pour une valeur donnée de n, le point $(x_1, x_2)$ qui minimise la variance de $\bar{y}_\omega$ doit être déterminé numériquement à l'aide d'un ordinateur.

## APPLICATIONS

Comme le programme informatique permettant d'effectuer l'analyse décrite dans la section précédente est encore au stade de développement, on a utilisé une méthode empirique simple pour étudier les effets du plan d'échantillonnage sur l'estimation de la moyenne pondérée en fonction de la surface du lac. On y parvient en supposant qu'il existe un plan d'échantillonnage permettant de représenter la «vraie» population. Le vecteur estimé $\hat{\underline{\alpha}}$ est donc considéré comme le «vrai» vecteur des paramètres de la population. Plusieurs plans d'échantillonnage sont ensuite générés à partir de cette population. Ainsi, le premier plan consiste à retenir une station sur deux; ce plan représente donc la moitié de la population. Le deuxième plan retient une station sur trois, etc. On présente à la figure 84 la variance de la température $\bar{y}_\omega$ en fonction du nombre de stations d'échantillonnage lorsque l'on considère les missions de 1968 et 1977 comme représentant la population réelle. La date de chaque mission est indiquée sur la courbe de variance correspondante. On note que la variance décroit sensiblement quand les plans d'échantillonnage comportent moins de 20 stations. La variance décroit très lentement au-dessus de 20 stations. Cela porte à croire qu'entre 20 et 25 stations suffisent pour l'estimation de la température moyenne pondérée en fonction de la surface. On en arrive à la même conclusion pour la teneur en chlorophylle $\underline{a}$ (figure 85). De plus, on présente à la figure 86 l'écart en pourcentage entre la moyenne pondérée en

fonction de la surface $\bar{y}_{\omega}$ et la moyenne pondérée «vraie» (basée sur toutes les stations échantillonnées pendant la mission) de la teneur en chlorophylle *a*. L'écart demeure inférieur à 7.5 % quand un plan d'échantillonnage de 25 stations est utilisé pour contrôler le lac.

Tableau 66. Moyenne, moyenne pondérée, variance, efficacité et nombre de stations pour chaque mission

| Date | En fonction de la surface Moyenne, $\bar{r}_a$ | Moyenne pondérée, $\bar{r}_w$ | Variance Réelle | Variance Optimale | Efficacité | Nombre de stations Réel n | Nombre de stations Estimé $\tilde{n}$ |
|---|---|---|---|---|---|---|---|
| 17-24 mai | 6.581 | 6.848 | 0.1239 | 0.0480 | 0.387 | 62 | 24.02 |
| 29 juil.-3 août | 13.771 | 11.712 | 0.1888 | 0.1362 | 0.720 | 61 | 44.00 |
| 31 août-3 sept. | 5.500 | 6.010 | 0.2801 | 0.1382 | 0.491 | 28 | 13.80 |
| 4-10 nov. | 5.000 | 3.730 | 0.0526 | 0.0305 | 0.580 | 55 | 32.00 |

Tableau 67. Affectation optimale des stations d'échantillonnage du lac Erié

| | 17 - 24 mai | | | | 29 juil. - 3 août | | | | 31 août - 3 sept. | | | | 4 - 10 nov. | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Moyenne | Nombre de stations | | | Moyenne | Nombre de stations | | | Moyenne | Nombre de stations | | | Moyenne | Nombre de stations | | |
| Zone | $\bar{r}$ | n | $\hat{n}$ | $\tilde{n}$ | $\bar{r}$ | n | $\hat{n}$ | $\tilde{n}$ | $\bar{r}$ | n | $\hat{n}$ | $\tilde{n}$ | $\bar{r}$ | n | $\hat{n}$ | $\tilde{n}$ |
| I | 0.566 | 43 | 20.46 | 7.92 | 0.750 | 24 | 6.10 | 4.4 | 1.318 | 22 | 12.0 | 6.0 | 0.306 | 36 | 17.6 | 10.2 |
| II | 7.833 | 13 | 11.16 | 4.32 | 9.684 | 19 | 30.50 | 22.0 | 13.600 | 5 | 10.8 | 5.4 | 5.700 | 10 | 12.0 | 7.0 |
| III | 35.500 | 4 | 22.94 | 8.88 | 25.000 | 10 | 11.00 | 7.9 | 38.000 | 1 | 5.2 | 2.6 | 17.140 | 7 | 19.5 | 11.4 |
| IV | 72.667 | 2 | 7.44 | 2.88 | 48.500 | 8 | 13.40 | 9.7 | | | | | 44.000 | 2 | 5.9 | 3.4 |

Tableau 68. Ecart relatif et taille de l'échantillon correspondante permettant de déceler des différences au niveau de signification de 5 %

| 17 - 24 mai | | 20 juil. - 3 août | | 31 août - 3 sept. | | 4 - 10 nov. | |
|---|---|---|---|---|---|---|---|
| $d/\bar{r}$ | n | $d/\bar{r}$ | n | $d/\bar{r}$ | n | $d/\bar{r}$ | n |
| 0.001 | 114 325 | 0.001 | 319 102 | 0.002 | 148 635 | 0.002 | 64 424 |
| 0.003 | 18 851 | 0.002 | 79 776 | 0.003 | 37 158 | 0.005 | 16 106 |
| 0.004 | 12 703 | 0.004 | 12 764 | 0.005 | 16 515 | 0.008 | 7 158 |
| 0.006 | 7 145 | 0.009 | 3 191 | 0.007 | 9 289 | 0.011 | 4 026 |
| 0.007 | 4 573 | 0.013 | 1 418 | 0.008 | 5 945 | 0.013 | 2 577 |
| 0.015 | 1 143 | 0.017 | 797 | 0.017 | 1 486 | 0.027 | 644 |
| 0.022 | 508 | 0.026 | 354 | 0.025 | 660 | 0.040 | 286 |
| 0.029 | 285 | 0.034 | 199 | 0.033 | 371 | 0.054 | 161 |
| 0.044 | 127 | 0.043 | 127 | 0.050 | 165 | 0.080 | 71 |
| 0.058 | 71 | 0.051 | 88 | 0.067 | 93 | 0.107 | 40 |
| 0.073 | 45 | 0.060 | 65 | 0.083 | 59 | 0.134 | 25 |
| 0.088 | 31 | 0.068 | 50 | 0.100 | 41 | | |
| 0.102 | 23 | 0.077 | 39 | 0.116 | 30 | | |
| 0.117 | 18 | 0.085 | 32 | 0.133 | 23 | | |

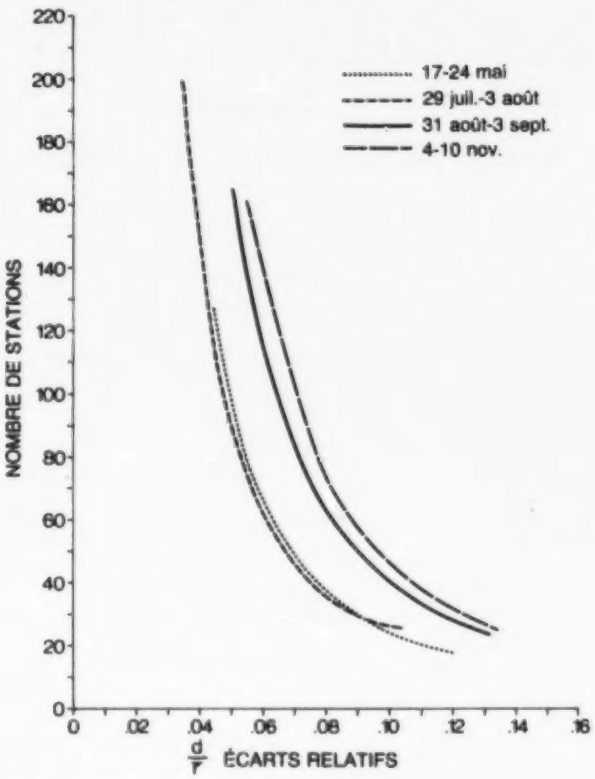



**Figure 83. Nombre de stations nécessaire pour déceler une différence dans la moyenne au niveau de signification de 5 % en fonction des écarts relatifs.**

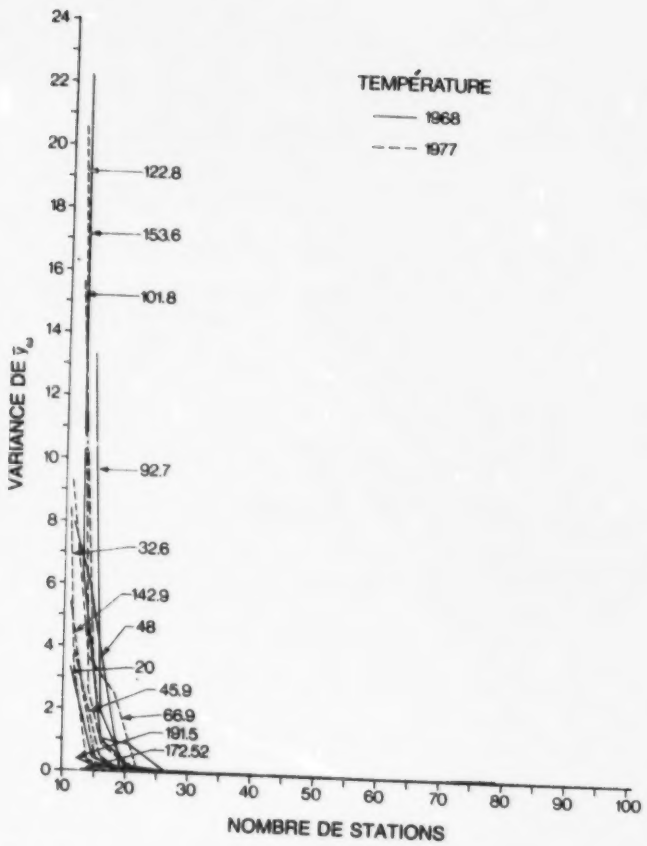



**Figure 84.** Variance de la température moyenne de surface en fonction du nombre de stations.

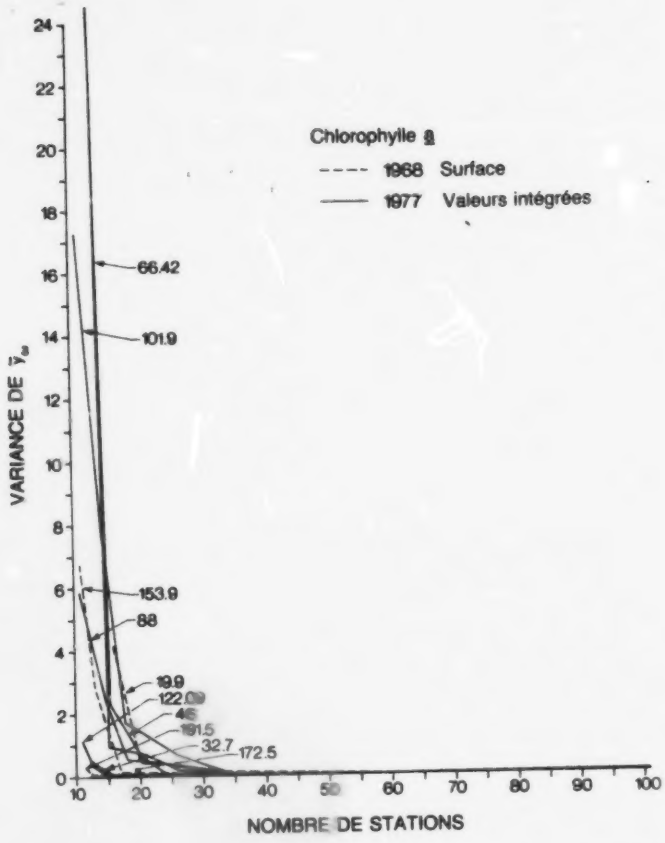



**Figure 85. Variance de la chlorophylle *a* en fonction du nombre de stations.**

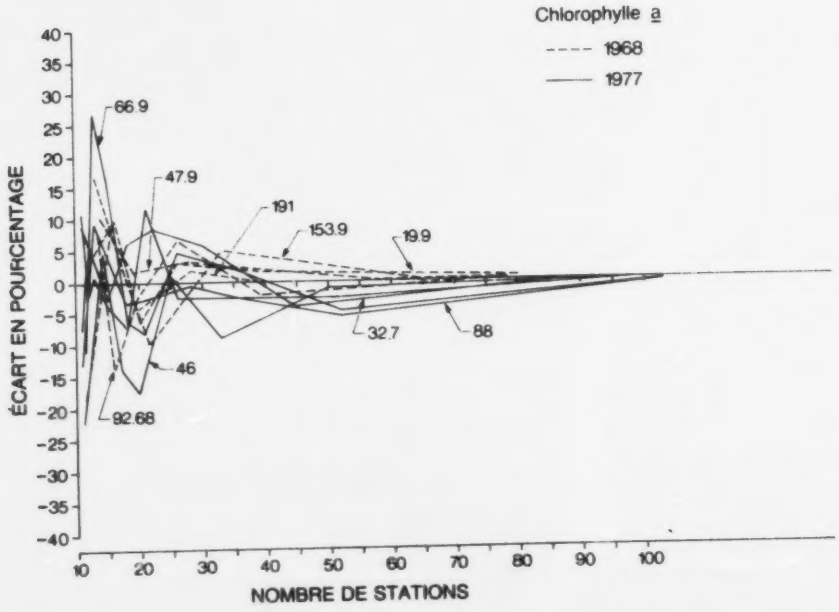



**Figure 86. Déviation en pourcentage de la chlorophylle *a* moyenne pour divers formats d'échantillons.**

# Références


Aitchison, J. et J.A.C. Brown. 1957. The Lognormal Distribution. Cambridge University Press, Cambridge, 176 p.

Anderson, J.E. 1982. Analysis of dissolved oxygen in Lake Erie (1967-1980) from measured data using cluster analysis. Thèse de maîtrise, Département de Génie, Université de Waterloo, Waterloo (Ontario).

Anderson, J.E., A.H. El-Shaarawi, S.R. Esterby et T.E. Unny. 1984. Variabilité spatiale et temporelle de l'oxygène dissous dans le lac Erié. Chapitre 4 de ce rapport.

Baldwin, N.S. et R.W. Saalfield. 1962. Great Lakes Fish. Comm. Tech. Rep. 3, plus supplément, 166 p.

Ball, G.H. et D.J. Hall. 1967. A clustering technique for summarizing multivariate data. Behav. Sci. 12:153-157.

Barica, J. 1982. Lake Erie depletion controversy. J. Great Lakes Res. 8(4):719-722.

Beeton, A.M. 1963. Limnological survey of Lake Erie, 1959 and 1960. Great Lakes Fish. Comm. Tech. Rep. 6, 35 p.

Beeton, A.M. 1965. Eutrophication of the St. Lawrence Great Lakes. Limnol. Oceanogr. 10:249-254.

Bierman, V.J., Jr. 1980. A comparison of models developed for phosphorus management in the Great Lakes, p. 235-255. Dans Phosphorus Management Strategies for Lakes, préparé par R.C. Loehr, C.S. Martin et W. Rast, Ann Arbor Sci., Ann Arbor, Mich., 490 p.

Birge, E.A. 1898. Plankton studies on Lake Mendota II. The Crustacea of the plankton from July, 1894, to December, 1896. Trans. Wis. Acad. Sci.Arts Lett. 11:274-448.

Birge, E.A. 1904. The thermocline and its biological significance. Trans. Am. Microsc. Soc. 25:5-33.

Birge, E.A. 1906. Gases dissolved in the waters of Wisconsin lakes. Trans. Am. Fish. Soc. 1906:143-163.

Birge, E.A. 1910. An unregarded factor in lake temperature. Trans. Wis. Acad. Sci. Arts Lett. 16(2):989-1016.

Birge, E.A. 1916. The work of the wind in warming a lake. Trans. Wis. Acad. Sci. Arts Lett. 18:341-391.

Birge, E.A. et C. Juday. 1919. Further limnological observations on the Finger Lakes of New York. Bull. U.S. Bur. Fish. 37.

Birge, E.A. et C. Juday. 1921. The inland waters of Wisconsin. The dissolved gases of the water and their biological significance. Wis. Geol. Nat. Hist. Surv. Bull. XXII.

Brönsted, I.N. et C. Wesenberg-Lund. 1912. Chemisch-physikalische Untersuchungen der dänischen Gewässer nebst Bemerkungen über ihre Bedeutung für unsere Auffassung der Temporalvariationen. Int. Rev. Hydrobiol. 4:251-290.

Brown, R.L., J. Durbin et J.M. Evans. 1975. Techniques for testing the constancy of regression relationships over time. J.R. Statist. Soc., Ser. B, 37, 149-192.

Brownlee, K.A. 1965. Statistical Theory and Methodology in Science and Engineering. John Wiley and Sons, Inc., New York, 590 p.

Burkholder, J.J. 1960. A survey of the microplankton of Lake Erie. U.S. Fish Wildl. Serv. Spec. Sci. Fish. Rep. No. 334, Washington, D.C., p. 123-144.

Burns, N.M. 1976a. Oxygen depletion in the Central and Eastern basins of Lake Erie, 1970. J. Fish. Res. Board Can. 33:512-519.

Burns, N.M. 1976b. Temperature, oxygen, and nutrient distribution patterns in Lake Erie, 1970. J. Fish. Res. Board Can. 33:485-511.

Burns, N.M. et F. Rosa. 1981. Oxygen depletion rates in the hypolimnion of central and eastern Lake Erie. A new approach indicates change. Ebauche inédite, présentée à l'atelier sur les controverses sur l'appauvrissement en oxygène, les 2 et 3 décembre 1981, Centre canadien des eaux intérieures, Burlington (Ontario).

Burns, N.M. et C. Ross. 1972a. Oxygen-nutrient relationships within the Central Basin of Lake Erie, p. 85-119. Dans Project Hypo, préparé par N.M. Burns et C. Ross, Centre canadien des eaux intérieures, p. 6, U.S. EPA Tech. Rep. TS-05-71-208-24, 182 p.

Burns, N.M. et C. Ross. 1972b. Project Hypo - Discussion on Findings, p. 120-126. Dans Project Hypo, préparé par N.M. Burns et C. Ross, Centre canadien des eaux intérieures, p. 6, U.S. EPA Tech. Rep. TS-05-71-208-24, 182 p.

Carr, J.F. 1962. Dissolved oxygen in Lake Erie, past and present. Great Lakes Research Division, University of Michigan, Pub. n° 9:1-14.

Carr, J.F. et J.K. Hitunen. 1965. Changes in the bottom fauna of western Lake Erie from 1930 to 1961. Limnol. Oceanogr. 10:551-569.

Centre canadien des eaux intérieures. 1966-1976. Lac Erié. Rapports sur les données limnologiques. Centre canadien des eaux intérieures, Burlington (Ontario).

Chapra, S.C. 1980. Application of the phosphorus loading concept to the Great Lakes, p. 135-152. Dans Phosphorus Management Strategies for Lakes, préparé par R.C. Loehr, C.S. Martin et W. Rast, Ann Arbor Sci., Ann Arbor, Mich., 490 p.

Charlton, M.N. 1979. Raréfaction de l'oxygène hypolimnétique du centre du lac Erié: La situation a-t-elle changée? Etude n° 110, Sér. Sci. Direction générale des eaux intérieures, Environnement Canada, Burlington (Ontario) 25 p.

Charlton, M.N. 1980a. Oxygen depletion in Lake Erie: Has there been any change? Can. J. Fish. Aquat. Sci. 37:72-81.

Charlton, M.N. 1980b. Hypolimnion oxygen consumption in lakes: Discussion of productivity and morphometry effects. Can. J. Fish. Aquat. Sci. 37:1531-1539.

Commission mixte internationale. 1975. Great Lakes Water Quality, 1974. 3rd Annual Report of the Great Lakes Water Quality Board to the International Joint Commission, Windsor (Ontario) 170 p.

Commission mixte internationale. 1976a. Great Lakes Water Quality, 1975. 4th Annual Report of the Great Lakes Water Quality Board to the International Joint Commission, Windsor (Ontario) 72 p.

Commission mixte internationale. 1976b. Great Lakes Water Quality, 1975. Appendix B, Surveillance Subcommittee Report, 255 p.

Commission mixte internationale. 1977. Great Lakes Water Quality, 1976. Appendix B, Surveillance Subcommittee Report, 134 p.

Commission mixte internationale. 1978a. Great Lakes Water Quality, 1977. 6th Annual Report of the Great Lakes Water Quality Board to the International Joint Commission, Windsor (Ontario) 89 p.

Commission mixte internationale. 1978b. Great Lakes Water Quality, 1977. Appendix B, Surveillance Subcommittee Report, 110 p.

Commission mixte internationale. 1979a. Great Lakes Water Quality, 1978. 7th Annual Report of the Great Lakes Water Quality Board to the International Joint Commission, Windsor (Ontario), 110 p.

Commission mixte internationale. 1979b. Great Lakes Water Quality, 1978. Appendix B. Surveillance Subcommittee Report, 117 p.

Commission mixte internatonale. 1980a. 1980 Report on Great Lakes Water Quality. Appendix, 82 p.

Commission mixte internationale. 1980b. Great Lakes Water Quality, 1979. 8th Annual Report of the Great Lakes Water Quality Board to the International Joint Commission, Windsor (Ontario), 150 p.

Conseil international sur la pollution des eaux du lac Erié et Conseil international sur la pollution des eaux du lac Ontario et du fleuve Saint-Laurent. 1969. Report to the International Joint Commission, Vol. 1, Summary, 150 p.

Dambach, C.A. 1969. Changes in the biology of the lower Great Lakes. Bull. Buffalo Soc. Nat. Sci. 25:1-10.

Davies, C.C. 1964. Evidence for the eutrophication of Lake Erie from phytoplankton records. Limnol. Oceanogr. 9:275-283.

Davies, C.C. 1969. Plants in Lake Erie and Ontario, and changes of their numbers and kinds. Bull. Buffalo Soc. Nat. Sci. 25:18-41.

DiToro, D.M. 1980. The effect of phosphorus loadings on dissolved oxygen in Lake Erie, p. 191-206. Dans Phosphorus Management Strategies for Lakes, préparé par R.C. Loehr, C.S. Martin et W. Rast, Ann Arbor Sci., Ann Arbor, Mich., 490 p.

Dobson, H.H. et M. Gilbertson. 1971. Oxygen depletion in the hypolimnion of the Central Basin of Lake Erie, 1929 to 1970. Proc. Great Lakes Res. 14:743-748.

Duda, R. et P. Hart. 1973. Pattern Classification and Scene Analysis. John Wiley and Sons, Inc., New York.

Durbin, J. 1969. Tests for serial correlation in regression analysis based on the periodogram of least squares residuals. Biometrika, 56:1-15.

El-Shaarawi, A.H. 1982. Stratégie d'échantillonnage pour estimer la densité des bactéries dans les Grands lacs. Journal français d'hydrologie, 13:171-187.

El-Shaarawi, A.H. 1984a. Variations temporelles dans le lac Erié. Chapitre 3 de ce rapport.

El-Shaarawi, A.H. 1984b. Dissolved oxygen concentrations in lake Erie. 2. Un modèle statistique de l'oxygène dissous dans le bassin central du lac Erié. Chapitre 5 de ce rapport.

El-Shaarawi, A.H. et S.R. Esterby. 1981. Analyse de régression de la variation spatiale d'une caractéristique limnologique. Eau du Québec, 14(3).

El-Shaarawi, A.H. et R.E. Kwiatkowski. 1977. A model to describe the inherent spatial and temporal variability of parameters in Lake Ontario, 1974. J. Great Lakes Res. (3-4), p. 177-183.

El-Shaarawi, A.H. et W.O. Pipes. 1982. Enumeration and statistical inferences, Chapter 3. Dans Bacterial Indicators of Pollution, préparé par W.O. Pipes, CRC Press, Boca Raton, Floride, 174 p.

El-Shaarawi, A.H. et K.R. Shah. 1978. Statistical procedures for classification of a lake. Etude n° 86, Sér. Sci. Direction générale des eaux intérieures, Environnement Canada, Burlington (Ontario).

El-Shaarawi, A.H., S.R. Esterby et B.J. Dutka. 1981. Bacterial density water determined by Poisson or negative binomial distributions. Appl. Environ. Microbiol., p. 107-116.

Environnement Canada. 1979. Analytical Methods Manual. Direction de la qualité des eaux, Direction générale des eaux intérieures, Ottawa.

Environnement Canada. 1979. Références sur la qualité des eaux. Guide des paramètres de la qualité des eaux, supplément. Direction générale des eaux intérieures, Direction de la qualité des eaux, Ottawa.

Esterby, S.R. 1982. Fitting probability distributions to bacteriological data: considerations for regulations and guidelines. Journal français d'hydrologie, 13:189-203.

Esterby, S.R. et A.H. El-Shaarawi. 1981. Characterization of spatial and temporal variability of water quality parameters. Summaries of Conference Presentations, Environmentrics 81, Alexandria, Virginia, p. 156-157.

Fay, L.A. et C.E. Herdendorf. 1981. Lake Erie water quality: Assessment of 1980 open lake conditions and trends for the preceding decade. CLEAR Technical Report No. 219, Ohio State University, Center for Lake Erie Area Research, Columbus, Ohio.

Feller, W. 1957. An Introduction to Probability Theory and its Applications. John Wiley and Sons, Inc., New York, 509 p.

Fish, C.J. 1960. Limnological survey of eastern and central Lake Erie, 1928 to 1929. U.S. Fish Wildl. Serv. Spec. Sci. Fish Rep. No. 334, Washington, D.C., p. 198.

Fisher, R.A. 1941. The negative binomial distribution. Ann Eugen. 11: 182-187.

Fisher, R.A., H.G. Thornton et W.A. MacKenzie. 1922. The accuracy of the plating method of estimating the density of bacterial populations. Ann. Appl. Biol. 9:325-359.

Hamblin, P.F. 1971. Circulation and water movement in Lake Erie. Sci. Ser. No. 7, Energie, Mines et Ressources, Ottawa, 50 p.

Harris, G. et R.A. Vollenweider. 1982. Paleolimnological evidence of early eutrophication in Lake Erie. Can. J. Fish. Aquat. Sci. 39:618-626.

Hoppe-Seyler, F. 1895. Über die Verteilung absorbierter Gase im Wasser des Bodensees und ihre Beziehungen zu den in ihm lebenden Tierem und Pflanzen. Schr. Ver. Gesh. Bodensees. 24:39-48.

Hutchinson, G.E. 1957. A Treatise on Limnology. Vol. I, John Wiley and Sons, Inc., New York, 1115 p.

Hutchinson, G.E. 1967. A Treatise on Limnology. Vol. II. Introduction to lake biology and the limnoplankton. John Wiley and Sons, Inc., New York, 1115 p.

Kalbfleisch, J.D. et D.A. Sprott. 1973. Marginal and conditional likelihoods. Sankhya, Ser. A, 35: 311-328.

Lam, D.C.L., W.M. Schertzer et A.S. Fraser. 1983. Simulation des réponses du lac Erié aux variations de charge et aux variations climatiques du point de vue de la qualité de l'eau. Etude n° 134, Sér. Sci. Direction générale des eaux intérieures, Environnement Canada, Burlington (Ontario).

Leach, J.H. et S.J. Nepsy. 1976. The fish community in Lake Erie. J. Fish. Res. Board Can. 33:622-638.

Lee, G.F., R.A. Jones et W. Rast. 1980. Availability of phosphorus to phytoplankton and its implications for phosphorus management strategies, p. 259-308. Dans Phosphorus Management Strategies for Lakes, préparé par R.C. Loehr, C.S. Martin et W. Rast, Ann Arbor Sci., Ann Arbor, Mich., 490 p.

Lund, J.W.G. 1965. The ecology of the fresh water phytoplankton. Biol. Rev. 40:231-293.

Nicholls, K.H. et P.J. Dillon. 1978. An evaluation of phosphorus - chlorophyll-phytoplankton relationships for lakes. Int. Rev. Gesamten Hydrobiol. 63:141-159.

Northouse, R.A. et F.R. Fromm. 1976. Class: A nonparametric clustering algorithm. Pattern Recognition, 8:107-114.

Pipes, W.O., P. Ward et S.H. Ahn. 1977. Frequency distributions for coliform bacteria in water. J. Am. Water Works Assoc. 69:664-668.

Phosphorus Management Strategies Task Force. 1980. Phosphorus Management for the Great Lakes. Final Report, 129 p.

Rao, S.S. et B.K. Burnison. 1976. Bacterial distributions in Lake Erie (1967, 1970). J. Fish. Res. Board Can. 33:547-580.

Reitz, R.D. 1973. Distribution of phytoplankton and Coliform Bacteria in Lake Erie. Ohio Environmental Protection Agency, Division of Surveillance.

Richards, R.P. 1981. Historical trends in water chemistry in the U.S. nearshore, central basin, Lake Erie. Water Quality Laboratory Tech. Rep. Ser. No. 15, Heidelberg College, Tiffin, Ohio, 31 p.

Rockwell, D.C., C.J. Marion, M.F. Palmer, D.S. DeVault et R.J. Bowden. 1980. Environmental trends in Lake Michigan. Proceedings of the 1979 Conference on Phosphorus Management Strategies for Lakes, préparé par R.C. Loehr, C.S. Martin et W. Rast, Ann Arbor Sci., Ann Arbor, Mich., p. 91-134.

Rosa, F. et N.M. Burns. 1981. Oxygen depletion rates in the hypolimnion of central and eastern Lake Erie. A new approach indicates change. Présenté à l'atelier sur l'appauvrissement en oxygène (bassin central) les 2 et 3 déc. 1981, Centre canadien des eaux intérieures, Burlington (Ontario).

Ruttner, F. 1952. Fundamentals of Limnology. Traduit par D.G. Frey et F.E.J. Fry, Presses de l'Université de Toronto, Toronto, 295 p.

Santé et Bien-être Canada. 1979. Directives sur l'eau potable au Canada. Centre des publications du gouvernement canadien, Approvisionnements et Services Canada, Hull (Québec).

Simons, T.J. 1976. Continuous dynamical computations of water transports in Lake Erie for 1970. J. Fish. Res. Board Can. 33:371-384.

Slater, R.W. et G.E. Bangay. 1980. Action taken to control phophorus in the Great Lakes. Proceedings of the 1979 Conference on phosphorus Management Strategies for Lakes, préparé par R.C. Loehr, C.S. Martin et W. Rast, Ann Arbor Sci., Ann Arbor, Mich., p. 13-26.

Sly, P.G., 1976. Lake Erie and its basin. J. Fish. Res. Board Can. 33:355-370.

Smirnov, V.I. 1964. Linear Algebra. Pergamon Press, Londres, 324 p.

Snedecor, G.W. et W.G. Cochran. 1967. Statistical Methods. The Iowa State University Press, Ames, Iowa, 593 p.

Sprott, D.A. et J.G. Kalbfleisch. 1965. Use of the likelihood function in inference. Psychol. Bull. 64:15-22.

Thienemann, A. 1915. Physikalische und chemische untersuchungen in den Meeren der Eifel. Verh. naturh. Ver. preuss. Rheinl. 71:281-389.

Thienemann, A. 1928. Der Sauerstoff in eutrophen und oligotrophen Seen. Die Binnengewässer, Vol. 4. E. Schweizerbart'sche Verlagsbuchhandlung, Stuttgard, Allemagne, 175 p.

Thomann, R.V. et J.S. Segna. 1980. Dynamic phytoplankton-phosphorus model of Lake Ontario: ten year verification and simulations, p. 153-170. Dans Phosphorus Management Strategies for Lakes, préparé par R.C. Loehr, C.S. Martin et W. Rast, Ann Arbor Sci., Ann Arbor, Mich., 490 p.

Thomas, N.A., A. Robertson et W.C. Songzogni. 1980. Review of control objectives, new target loads and input controls, p. 61-90. Dans Phosphorus Management Strategies for Lakes, préparé par R.C. Loehr, C.S. Martin et W. Rast, Ann Arbor Sci., Ann Arbor, Mich., 490 p.

Vollenweider, R.A. et L.L. Janus. 1981. Statistical models for predicting hypolimnetic oxygen depletion rates. Présenté à l'atelier sur l'appauvrissement de l'oxygène (bassin central), les 2 et 3 déc. 1981, Centre canadien des eaux intérieures, Burlington (Ontario).

Vollenweider, R.A., W. Rast et J. Kerekes. 1980. The phosphorus loading concept and Great Lakes eutrophication, p. 207-234. Dans Phosphorus Management Strategies for Lakes, préparé par R.C. Loehr, C.S. Martin et W. Rast, Ann Arbor Sci., Ann Arbor, Mich., 490 p.

Wetzel, R.G. et G.E. Likens. 1979. Limnological Analysis. W.B. Saunders Co., Toronto.

Zar, H. 1980. Point source loads of phosphorus to the Great Lakes, p. 27-36. Dans Phosphorus Management Strategies for Lakes, préparé par R.C. Loehr, C.S. Martin et W. Rast, Ann Arbor Sci., Ann Arbor, Mich., 490 p.

# Annexe

# Examen de modèles de régression linéaire par analyse des résidus

# Examen de modèles de régression linéaire par analyse des résidus

## INTRODUCTION

On traite, dans la présente annexe de la pertinence des modèles de régression linéaire. L'examen porte sur l'analyse des résidus qui sont les écarts entre les valeurs observées et les valeurs estimées correspondantes. Plusieurs fonctions de résidus sont présentées et l'on explique comment elles servent à déceler la pertinence des modèles de régression.

## MODELES DE REGRESSION LINEAIRE

Soit $\underline{y}$ le vecteur colonne $(y_1, y_2 \ldots y_n)$ de n variables aléatoires généré pour le modèle linéaire

$$\underline{y} = A\underline{\theta} + \underline{\varepsilon} \qquad \text{(A-1)}$$

où A est la matrice d'ordre (n x k) et de rang complet k, $\underline{\theta}$ est un vecteur de k paramètres inconnus et $\underline{\varepsilon}$ est un vecteur de n variables aléatoires indépendantes distribuées normalement et impossible à observer, de moyenne $\underline{0}$ et de variance $\sigma^2$. Il est bien connu que la meilleure estimation linéaire non biaisée de $\theta$ est

$$\underline{\hat{\theta}} = (A'A)^{-1} A'\underline{y} \qquad \text{(A-2)}$$

où, pour toute matrice B, les matrices B' et $B^{-1}$ sont respectivement la transposée de B et l'inverse de B (si elle esiste). L'estimation de $\underline{y}$ est:

$$\hat{\underline{y}} = A\hat{\theta} \qquad \text{(A-3)}$$

En utilisant la relation A-2, on peut réécrire l'équation A-3 sous la forme:

$$\hat{\underline{y}} = P\underline{y} \qquad \text{(A-4)}$$

où $p = A(A'A)^{-1}A'$ est la matrice de projection. La matrice P est idempotente ($P^2 = P$) et de rang k. Le vecteur des résidus est:

$$\begin{aligned} \underline{R} &= \underline{y} - \hat{\underline{y}} \\ &= (I - P)\underline{y} \end{aligned} \qquad \text{(A-5)}$$

où I est la matrice unité. La somme des carrés résiduelle est:

$$\begin{aligned} S &= \underline{R}'\underline{R} \\ &= \underline{y}'(I - P)\underline{y} \end{aligned} \qquad \text{(A-6)}$$

La variance $\sigma^2$ est estimée par

$$\hat{\sigma}^2 = S/(n-k) \qquad \text{(A-7)}$$

## QUELQUES PROPRIETES DES RESIDUS

Le fait que $\hat{\underline{\theta}}$ et $\hat{\sigma}^2$ suffisent à estimer les paramètres $\theta$ et $\sigma^2$ montre que le vecteur des résidus contient toute l'information nécessaire pour vérifier la pertinence du modèle A-1, surtout l'hypothèse sur la distribution $\underline{\varepsilon}$. Il convient donc

d'étudier la relation entre $\underline{\varepsilon}$ et $\underline{R}$. Etant donné que I - P et k sont orthogonaux, on obtient

$$\underline{R} = (I - P)\underline{\varepsilon} \qquad \text{(A-8)}$$

$$\underline{R}'\underline{R} = \underline{\varepsilon}'(I - P)\underline{\varepsilon} \qquad \text{(A-9)}$$

Il est possible, à partir de A-8, d'exprimer les éléments de $\underline{R}$ sous la forme de n combinaisons linéaires des éléments de $\underline{\varepsilon}$. Le rang de I - P n'étant pas complet mais égal à (n - k), ces combinaisons linéaires ne sont pas uniques. Les hypothèses concernant $\underline{\varepsilon}$ permettent de poser

$$E(\underline{R}) = 0 \qquad \text{(A-10)}$$

$$\text{var}\,(\underline{R}) = (I - P)\,\sigma^2 \qquad \text{(A-11)}$$

R a donc une distribution normale à plusieurs variables, dont la moyenne et la matrice des variances-covariances sont, respectivement, présentées en A-10 et A-11. Les éléments de $\underline{R}$ sont donc corrélés, et cette corrélation dépend uniquement de la matrice A. Quand certaines conditions touchent les éléments de A et que la valeur de n est élevée et celle de k est faible, la corrélation entre les éléments de $\underline{R}$ est très faible et peut être négligée; les éléments de P sont alors nuls dans A-11. Le rapport $\underline{R}/\sigma$ est donc considéré comme presque normal, avec une moyenne $\underline{0}$ et une matrice des variances-covariances I.

Considérons, par exemple, le modèle

$$\underline{Y} = \underline{1}\theta + \underline{\varepsilon}$$

où $\underline{1}$ est un vecteur unité et $\theta$ est la moyenne de $\underline{y}$. Dans ce cas, k = 1 et P = 1 1'/n. Les éléments hors diagonale de P étant alors tous égaux à 1/n, la corrélation entre les éléments de $\underline{R}$ peut donc être négligée pour les valeurs élevées de n.

Une autre approche, particulièrement utile quand les valeurs de n sont moyennes ou faibles, consiste à transformer $\underline{R}$ en un ensemble de (n - k) résidus orthogonaux. La formule de Jacobi (Smirnov, 1964, p. 130) est un outil pratique pour effectuer cette transformation.

## FORMULE DE JACOBI

Supposons que $Q = \underline{x}'C\underline{x}$ ait une forme quadratique et que $\underline{Z} = C\underline{x}$. Définissons la matrice carrée $C_k$ comme la matrice formée des k premières rangées et des k premières colonnes de C; $M_k$ comme la matrice $C_k$ dont la dernière colonne est remplacée par $Z_k$, le vecteur de longueur k dont les éléments sont les k premières composantes de Z. Soit

$$\Delta_0 = 0,\ \Delta_1 = a_{11},\ \Delta_k = |C_k| \quad (k = 2...n)$$

$$X_1 = Z_1,\ X_k = |M_k|, \quad (k = 2,3...n)$$

où $\Delta_k$ et $X_k$ sont les déterminants de $C_k$ et $X_k$ respectivement. Si le rang de la matrice C est r et si les déterminants $\Delta_1$, $\Delta_2 ... \Delta_r$ sont nuls, la formule de Jacobi devient

$$Q = \underset{-}{x}'C\underset{-}{x} = \sum_{k=1}^{r} \frac{X_k^2}{\Delta_k \Delta_{k-1}}$$

où les formes linéaires $X_k$ (k = 1,2...r) sont linéairement indépendantes.

L'application de la formule de Jacobi donne la série suivante de résidus orthogonalisés:

$$w_i = \{y_i - \underline{a}_i' \hat{\underline{\theta}}_{i-1}\} / \sqrt{1 + \underline{a}_i' (A_{i-1}' A_{i-1})^{-1} \underline{a}_i} \quad (i = k+1, k+2 \ldots n)$$

où $\hat{\underline{\theta}}_i$ est l'estimation aux moindres carrés de $\underline{\theta}$ à partir des i premières observations, $\underline{a}'_i$ est la $i^{ième}$ rangée de A et $A_i$ est la matrice des i premières rangées de A.

La normalité de $\varepsilon_i$ peut ensuite être testée en testant la normalité de $w_i$. Si (n-k) est grand, l'histogramme des fréquences observées de $w_i$ peut alors faire l'objet d'un test de normalité par la méthode classique du $x^2$. Si (n-k) est petit, des méthodes graphiques, comme les diagrammes P - P ou Q - Q, permettent alors de déceler tout écart par rapport à la normalité. On utilise, pour tester si $E[\varepsilon_i] = 0$ et $Var[\varepsilon_i] = \sigma^2$, les méthodes (i) des résidus cumulées (RC) et (ii) des résidus au carré cumulés (RCC). L'application de ces méthodes est décrite plus bas.

## LE TEST RC

La statistique du test est définie comme:

$$C_t = \frac{1}{\sigma} \sum_{j=k+1}^{t} w_j \qquad t = k+1 \ldots n \qquad (A\text{-}12)$$

Dans le cas de l'hypothèse nulle,

$$H : E(\varepsilon_i) = 0 \qquad i = 1,2 \ldots n$$

l'ensemble des variables aléatoires $C_{k+1}$, $C_{k+2} \ldots C_n$ a une distribution presque normale à plusieurs variables, de moyenne $\underline{0}$. Les éléments de la matrice des variances-covariances de $C_t$ (t = k+1...n) sont donnés par

$$\text{var}(C_t) = 1 \qquad t = k+1 \ldots n$$

et

$$\text{cov}(C_t, C_{t+r}) = t-k$$

Brown, Durbin et Evans (1975) ont montré que les valeurs critiques $C_a(t)$ de $C_t$ sont déterminées par

$$C_a(t) = h_a \sqrt{n-k} + z \cdot h_a \cdot \frac{t-k}{\sqrt{n-k}} \qquad t=k+1 \ldots n$$

où $h_a$ est la racine de l'équation

$$1 - \Phi(3h_a) + \Phi(h_a)\, e^{-2h^2_a} = a/2$$

où $$\Phi(h_a) = \int_{-\infty}^{h_a} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

## LE TEST RCC

Brown, Durbin et Evans (1975) ont présenté une statistique apparentée pour tester la constance des variances de $\varepsilon_1, \varepsilon_2 \ldots \varepsilon_n$; elle est donnée par

$$CS_t = sw_t / sw_n \qquad t = k+1 \ldots n$$

où

$$sw_r = \sum_{j=k+1}^{r} w_j^2$$

Dans le cas de l'hypothèse nulle:

$$E(CS_t) = (t-k)/(n-k) \cdot t = k+1 \ldots n$$

les valeurs critiques $CS_\alpha(t)$ servent à évaluer en simultanéité le degré de signification des écarts des $CS_t$ par rapport à leurs valeurs prévues. Ces valeurs critiques sont présentées par Durbin (1969).